PITCH PERCEPTION MECHANISMS IN MARMOSETS

by
Xindong Song

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

December 2016

ABSTRACT

The mechanisms of pitch perception have been one of auditory neuroscience's central questions for over a century due to the importance of pitch in music and speech perception. Yet the evolutionary origins of pitch perception, and whether its underlying mechanisms are unique to humans, is unknown. For my dissertation, I have investigated the perceptual properties of pitch in marmoset monkeys.

One of the most well-known phenomena of pitch perception, that of the missing fundamental, has suggested that humans do not simply use the fundamental frequency component to perceive pitch from harmonic complex sounds but can actively infer the pitch from the higher overtones or harmonics. It has been suggested that several non-human species are also sensitive to the pitch of missing fundamental sounds. However, none of these demonstrations has shown this sensitivity to pitch with a precision below three semitones. For humans to perceive Western music melodies, a precision of at least one semitone is necessary. The first step of my thesis was to confirm that marmoset monkeys can also perceive pitch through missing fundamental sounds with at least one semitone precision, using a behavioral generalization paradigm.

The next step of my thesis is to determine the mechanisms behind pitch perception of harmonic complex sounds in marmosets. It has been shown that humans hear the pitch of harmonic sounds through spectral or temporal features. Over the last century of human psychophysics research, three primary features of human pitch perception mechanisms have been described: (1) Lower resolved harmonics have a stronger pitch strength compared to a pure tone at the fundamental frequency (F0) and

also to higher unresolved harmonics; (2) pitch of resolved harmonics is sensitive to the quality of spectral harmonicity; (3) pitch of unresolved harmonics is sensitive to the salience of temporal envelope cues. Among these features, the first two have never been demonstrated in any other species besides humans. For this part of my thesis, I provided evidence that the marmoset, a highly vocal New World monkey species separated from humans by about 30 to 40 million years, exhibits all three primary features of pitch perception mechanisms as found in humans.

Combined with previous neurophysiological findings of a specialized pitch processing region in both marmoset and human auditory cortex, these findings suggest that the mechanisms for pitch perception, which have long been thought unique to humans, may have originated early in primate evolution, before the separation of New World and Old World primates.

Advisor:	Dr. Xiaoqin Wang Professor Biomedical Engineering	Reader:	Dr. Eric Young Professor Biomedical Engineering

ACKNOWLEDGEMENTS

First, I want to thank my advisor, Dr. Xiaoqin Wang. He encourages students to pursue ambitious and systematic work and gives us incredible amount of freedom to find our own interest and passion. He is such an energetic person that whenever I felt frustrated, I can always found some energy from talking to him. He once said working in this lab is like swimming in the ocean, but not like in a swimming pool. I did enjoy my years of fighting in the open water. It was thrilling but very fun. I also appreciate his insight and vision, which has established the marmoset neuroscience platform that all lab members benefit from.

I also want to thank all my thesis committee members. Dr. Eric Young not only taught me a lot through his course and journal club but also provided me many very sharp inputs and technical critiques of this thesis. Dr. Xingde Li generously provided his lab space and very precious equipment to us to test our first several two-photon imaging prototypes. Dr. Ed Connor and Dr. Kristina Nielsen also had many helpful discussions with us and helped us with their great expertise.

Dr. Michael Osmanski was very instrumental in this work. He kindly taught me everything about marmoset psychoacoustics without any reservation. He built chamber 4 and 5, from where all data of this work were collected. I enjoyed every moment I spent with him and had a very good time working and collaborating with him.

I owe my thanks to Yueqi Guo and Jian Lu. They helped me in collecting part of the data of this work. Yueqi also helped me throughout the development of two-photon surgical procedures and was very patient to very stressful me. All our animal work cannot be done without the support from our excellent technicians, Jenny Estes, Nate Sotuyo, Shanequa Smith, and Alexandra Prado.

I also want to thank all other members in the Wang lab. Lingyun Zhao was sitting in the same office with me for many years and shared many emotional moments with me. Evan Remington trained me how to do single unit recording. Yi Zhou, Luke Johnson, Marcus Jeschke, Lei Feng, and Sabyasachi Roy gave me many critical suggestions during my initial years. Darik Gamble, Kai-yuen Lim and Lingyun shared many working weekends with me when there's barely anybody else in the building. Yunyan Wang helped us designed two journal covers and was the go-to person in the lab. Lixia, Seth also provided me warm helps whenever I need from them.

Life would be much tougher without the support I got from my wonderful roommates of 222 E University Parkway in Baltimore, including but not limited to Alan, Lisa, Yun-ching, Yunke, Ariel, Amy, Lin and Dan. I especially want to thank Dan, who helped and supported me tremendously through my toughest years.

Finally, and most importantly, thanks to my parents for their unconditional love, support, and sacrifice for me to pursue my dream.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. What Is Pitch?

When we think about pitch perception, we may first think about music. A sequential presentation of pitch notes in semitone steps may form a melody while the simultaneous presentation of several pitch notes may produce a harmony. It is hard to imagine what music would sound like without pitch perception. It is also hard to imagine how tonal languages would work without pitch, as different tones or different pitches carry not only emotional information but also basic semantic information. Before tapping into details, the first question that must be answered is, what is pitch?

As its natural link to music, pitch can be defined in relation to music. The American Standards Association has suggested: “pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale” [ASA 1960]. Per this definition, pitch can be quantified in semitones on the music scale. As on the music scale, one octave is a doubling in temporal periodicity of the sound. One octave can be divided logarithmically into twelve equal steps as semitones. Thus, one semitone is about six percent of temporal periodicity change in the sound.

Alternatively, pitch can be defined not directly related to music. In the case of the definition from the American National Standards Institute: “pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high” [ANSI 1994].

As emphasized in both cases, pitch, unlike frequency, is an attribute of auditory sensation, but not a physical description of a sound. Although, for pure tones, which are

composed of a single spectral component, pitch and frequency are generally similar. For harmonic complex sounds, which are usually composed of a series of frequencies on integer multiples of a fundamental frequency (F0), the pitch is typically related to their F0, which also has the same temporal periodicity as the complex sound's temporal envelope periodicity. The details of the pitch of a harmonic complex sound will be discussed later.

Pitch perception may rank a sound as "low" or "high", either on music scale or not directly on the music scale. However, most would agree that the production of melodies is *sufficient* to prove that a sound can evoke a pitch. This requires pitch discrimination to be at least precise enough for a one semitone difference, which is the smallest step on the Western music scale.

1.2. Pitch of Pure Tones and Pitch of Complex Sounds

Many naturally oscillating objects generate a spectrum composed of a series of harmonically related frequency components. A simple example is an ideal string fixed at both ends as illustrated in figure 1.1. When a perturbation is introduced, the string can oscillate in different modes. The simplest mode generates a single oscillation frequency, and the other modes generate oscillation frequencies which are integer multiples of the simplest mode. The oscillation frequency of the simplest mode is thus called the fundamental frequency (F0) while the oscillation frequencies of other oscillation modes are called harmonics. The second harmonic (H2) bears a frequency twice that of the F0, the third harmonic (H3) bears a frequency which is three times that of the F0, and so on. Although a pure tone is sufficient to evoke pitch perception, a harmonic complex sound

is more commonly seen in human speech, animal vocalizations, and music. The pitch perceived from the complex sound is the same pitch as perceived from a pure tone at the F0 alone.

1.2.1 Missing fundamental pitch perception

Since a harmonic complex sound and a pure tone at its F0 alone can evoke the same pitch, it is natural to think the F0 component may serve as the key to perceiving pitch from a harmonic complex sound. Indeed, the very earliest pitch studies, just several decades after Fourier series were invented, proposed that the pitch of a complex sound is derived from the lowest harmonic, which is the F0 component [Ohm 1843, Helmholtz 1863]. Although, around the same time Seebeck showed that when there is little power left at the F0 in a harmonic complex sound, the pitch percept is still salient [Seebeck 1841]. This result started to cast doubt on the idea that the F0 component is not necessary for a pitch to be perceived from harmonic complex sounds.

After a century, efforts were made to completely remove the F0 component from the sound. The pitch was still intact [Schouten 1938]. One possible explanation of this phenomenon is, by theory, if the auditory system is nonlinear enough, it may re-introduce the F0 component back onto the basilar membrane by nonlinear interactions between adjacent higher harmonics in the cochlea [Thurlow and Small 1955]. Licklider suggested that this is not the case, however. He designed an experiment with a broadband low-frequency noise to mask any potential non-linear distortion product at the F0 in the auditory system, and the pitch percept was still intact [Licklider 1956]. Since the subject

cannot use the F0 nonlinear distortion product to infer the pitch, the pitch must be derived from higher harmonics without the F0 component.

Since the F0 component does not need to be present, either directly through sound delivery or through nonlinear distortions in the cochlea, for a pitch to be perceived from a harmonic complex sound, this phenomenon is thus called the “missing fundamental” and is one of the most noteworthy features of pitch perception [Plack 2005]. For instance, a harmonic complex sound composed with of components of 200, 300, and 400 Hz, but without a 100Hz component, can evoke the same pitch of a 100 Hz pure tone.

Missing fundamental pitch perception can effectively explain why we perceive pitch saliently through telephone service without difficulty. As a standard telephone line transmits a bandpass power between ~300 Hz and ~3000 Hz and cuts out lower frequencies, which normally includes the F0s of human voices.

Since the pitch from a missing fundamental sound cannot be derived simply by the lowest harmonic presented or introduced by the system’s nonlinearity, the auditory system must derive the pitch purely from higher harmonics by some other mechanisms, which are introduced below.

1.3. Candidate Mechanisms for Human Complex Sound Pitch Perception

In this part, candidate mechanisms for human complex sound pitch perception are introduced. This topic is also discussed in the content of F0 discrimination thresholds in the chapter 4.3.

1.3.1 Peripheral limitation of harmonic resolvability

The auditory system is hierarchically organized. A sound must pass through peripheral stages before being converted into an electrical neural code and entering the central nervous system.

After the outer and middle ear's transmission, the sound enters the inner ear, the cochlea. The cochlea spreads different frequencies along a tonotopic axis on the basilar membrane. Low frequencies travel more towards the distant, apical end, and high frequencies stay more to the basal end. This frequency spread does not occur in a linear fashion but roughly in a logarithmic fashion. The frequency resolution on the basilar membrane is not perfect. Each single frequency occupies a certain length along the tonotopic axis. For any given frequency, the frequency range along the basilar membrane over which it cannot be separated from other frequencies is roughly in a constant proportion to the frequency itself. Thus, for a harmonic complex sound, only the first five to ten harmonics are well separated on the basilar membrane and can be called resolved. The other way to describe this resolvability limitation is with the reference to a bank of auditory filters on the basilar membrane that separate different frequencies into different auditory channels. Each channel has a certain bandwidth and the bandwidths are approximately proportional to the center frequencies. For a harmonic complex sound, only the first five to ten harmonics are well separated into different auditory channels. When frequency increases, the bandwidth of the auditory filter increases as well. Thus, more harmonics enter the same auditory filter, and cannot be spectrally resolved. Those harmonics are so-called unresolved harmonics. The quantitative details and further analysis of this resolvability limitation is discussed in chapter 2.5.

Across a range of F0s (on the Y axis), the number of harmonics (on the X axis) that can be resolved is demonstrated in the figure 1.2 (A), as the line labeled as "presumable upper boundary of resolved harmonics". This number is within the range of five to ten across a wide range of F0s. Thus, the line looks roughly vertical in the plot. The criterion to define this boundary is discussed in chapter 2.5.2.

One thing noticeable in the plot is that each F0 has a different total number of harmonics available. A lower F0 has a larger number of harmonics available, whereas, a higher F0 only has a few number of harmonics available. Assuming an adult human's upper hearing boundary is around 16 kHz. A 100 Hz F0 has potentially 160 harmonics available, whereas a 3.2 kHz F0 can only have 5 harmonics available. Thus, a 16-kHz frequency forms a straight line going from the upper-left corner to the lower-right corner in the plot. A harmonic can only contribute when it can be heard. Thus, a pitch theory can only exist below this line.

1.3.2 Spectral theories

One of the potential mechanisms for harmonic complex sound pitch perception is through spectral analysis. If the incoming frequency components can be matched to an internal spectral harmonic template, then a corresponding pitch can be perceived [Goldstein 1973, Terhardt 1974, Shamma and Klein 2000]. Since spectral harmonic templates require each frequency component to be separable and individually represented in the system before being processed by the templates, they only work on resolved harmonics. As introduced above, not every harmonic can be resolved and individually

represented in the central system. Harmonic spectral templates are thus only able to work within the green area indicated in figure 1.2 (B).

1.3.3 Temporal theories

Frequency and temporal period are mathematically reciprocal to each other. A harmonic complex sound typically also has a temporal envelope periodicity which equals to F0. Through unresolved harmonics, since several adjacent harmonics fall into the same auditory filter and interact with each other, a temporal beating is generated on the temporal envelope, which bears a periodicity of F0. This temporal envelope cue can serve as a basis to infer the pitch from the sound [Schouten 1938] on unresolved harmonics, shown in the blue area indicated in figure 1.2 (C).

Moreover, temporal information is not only available through temporal envelope cues on unresolved harmonics but is also theoretically available through temporal fine structures within each auditory channel. For example, if a resolved harmonic has a frequency that is lower than the presumable phase locking limit (~5 kHz) of auditory nerves, the spikes on that auditory nerve may carry temporal fine structure that is phase locked to that frequency. Both temporal fine structures and temporal envelope cues can be extracted on a single channel by autocorrelation. The pitch percept can be represented by an aggregated periodicity function - the summary autocorrelation function (SACF) across all available channels [Meddis and O'Mard 1997, 2006]. This type of temporal theory expands the potential existence region of the temporal envelope theory from the unresolved side into the resolved side, as the green area indicated in figure 1.2 (D). This

autocorrelation-based temporal pitch theory has been proposed to solely explain pitch perception [Meddis and O'Mard 1997].

However, recent results indicate that, musical melodies can still be discriminable and pitch can still be heard for harmonic sounds that are presumably composed of resolved harmonics beyond the phase locking limit and thus are outside the possible existence region of any autocorrelation-based temporal pitch theory [Oxenham et al 2011, Oxenham and Micheyl 2013]. Since they are resolved harmonics of frequencies that are above the presumable phase locking limit (~5 kHz), theoretically, neither temporal fine structures nor temporal envelope cues is available.

There might be two possible explanations. One is that the human phase locking frequency limit is higher than that measured in other mammals. Although this is theoretically possible, it is noticeable that recent auditory nerve recordings from macaque monkeys [Michelet et al 2011] show a very similar phase locking limit compared with other mammals. Another possibility is that there is no unitary pitch theory to cover all details of pitch perception. Pitch perception, more likely, is composed of at least two mechanisms, as illustrated in figure 1.2 (E). These different pitch mechanisms must then be finally unified together in the brain to form a unitary pitch percept.

1.3.4 Summary

To sum up, for a wide range of F0s, about the first five to ten harmonics are well separated in the human auditory system. A spectral analysis using harmonic template matching can extract the pitch from the sound through these resolved harmonics. For the upper unresolved harmonics, a temporal analysis using temporal envelope cues can

extract the pitch from the sound. There is evidence that so far, a unitary temporal theory cannot fully cover the existence region of pitch perception. It is most likely the case that the pitch of a harmonic complex sound is derived from at least two distinct mechanisms, using either a spectral analysis based on harmonic templates, or a temporal analysis based mainly on temporal envelope cues.

1.4. Questions and Specific Aims

Recent studies have discovered pitch-selective neurons in a region of marmoset auditory cortex homologous to the pitch center found in human auditory cortex [Bendor and Wang 2005, Norman-Haignere et al 2013, Penagos et al 2004]. These neurons can encode a sound's pitch using either spectral or temporal cues, depending on the fundamental frequency and spectral resolvability [Bendor et al., 2012]. This neuronal substrate may serve as the key node of functional circuit for pitch processing.

These findings have led me to hypothesize that marmoset monkeys may also perceive pitch through human-like pitch perception mechanisms, as described above. The specific questions and aims of the current dissertation are listed below.

1.4.1 Do marmosets have precise missing fundamental pitch perception?

The first major question I want to ask is whether marmosets also have missing fundamental perception just as humans. There has been a number of animal behavioral studies showing several different species, ranging from fish to birds to mammals, possess missing fundamental perception [Fay 2005, Cynx and Shapiro 1986, Okanoya 2000, Heffner and Whitfield 1976, Chung and Colavita 1976, Tomlinson and Schwarz 1988].

However, none of these studies have shown evidence that a non-human species has missing fundamental perception with a precision of one semitone, which is the smallest pitch step on the Western music scale. So, if marmosets do have missing fundamental pitch perception, a further question is whether it is precise enough to potentially allow for fine melody discrimination. These questions are addressed in chapter 3 of this dissertation.

1.4.2 Do marmosets share the same pitch perception mechanisms of complex sounds with humans?

In humans, for complex sounds, pitch can be perceived through either spectral harmonicity analysis on resolved harmonics or temporal envelope analysis on unresolved harmonics. However, so far, nonhuman species have only shown sensitivity to temporal pitch but not to spectral pitch [Shofner and Chaney 2013, Joly et al 2014]. Our next major question is whether marmosets also have spectral harmonicity based pitch. If the answer is yes, a further question is, whether it is also the dominant pitch perception mechanism just as it is in humans [Plack 2005]. These questions are addressed in chapter 4 of this dissertation.

1.4.3 What's next?

If the answers of the last two questions I asked are both positive, how can we further target and understand the functional neural circuit underlying human-like pitch perception? In chapter 5, several further topics and some technical developments are discussed.

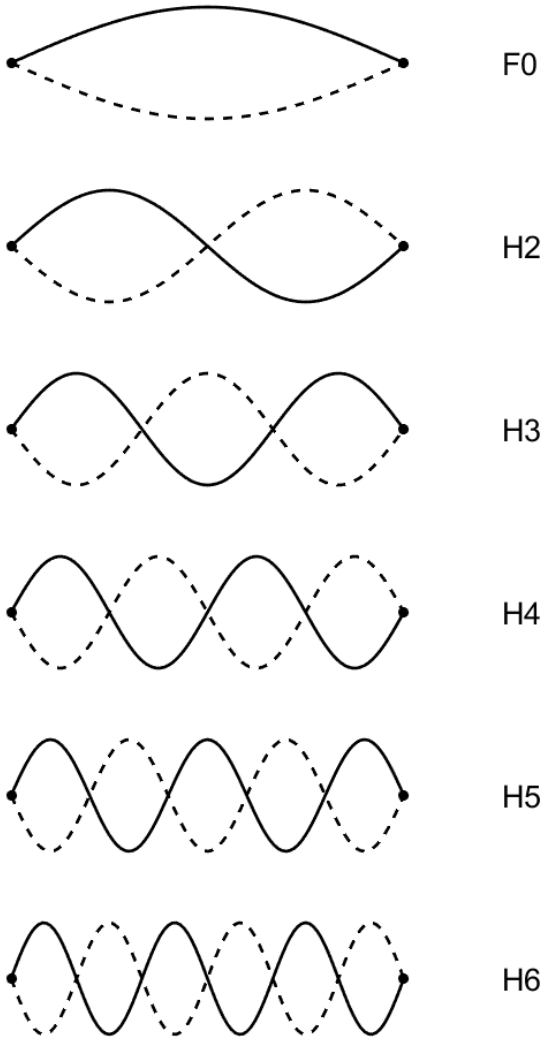


Figure 1.1. Oscillation modes of a string fixed at both ends

Oscillation modes of a string, fixed at both ends. The uppermost panel shows the simplest oscillation mode which bears an oscillation frequency as F0. The second oscillation mode bears an oscillation frequency H2 = 2 x F0. The third mode bears H3 = 3 x F0, and so far, so on.

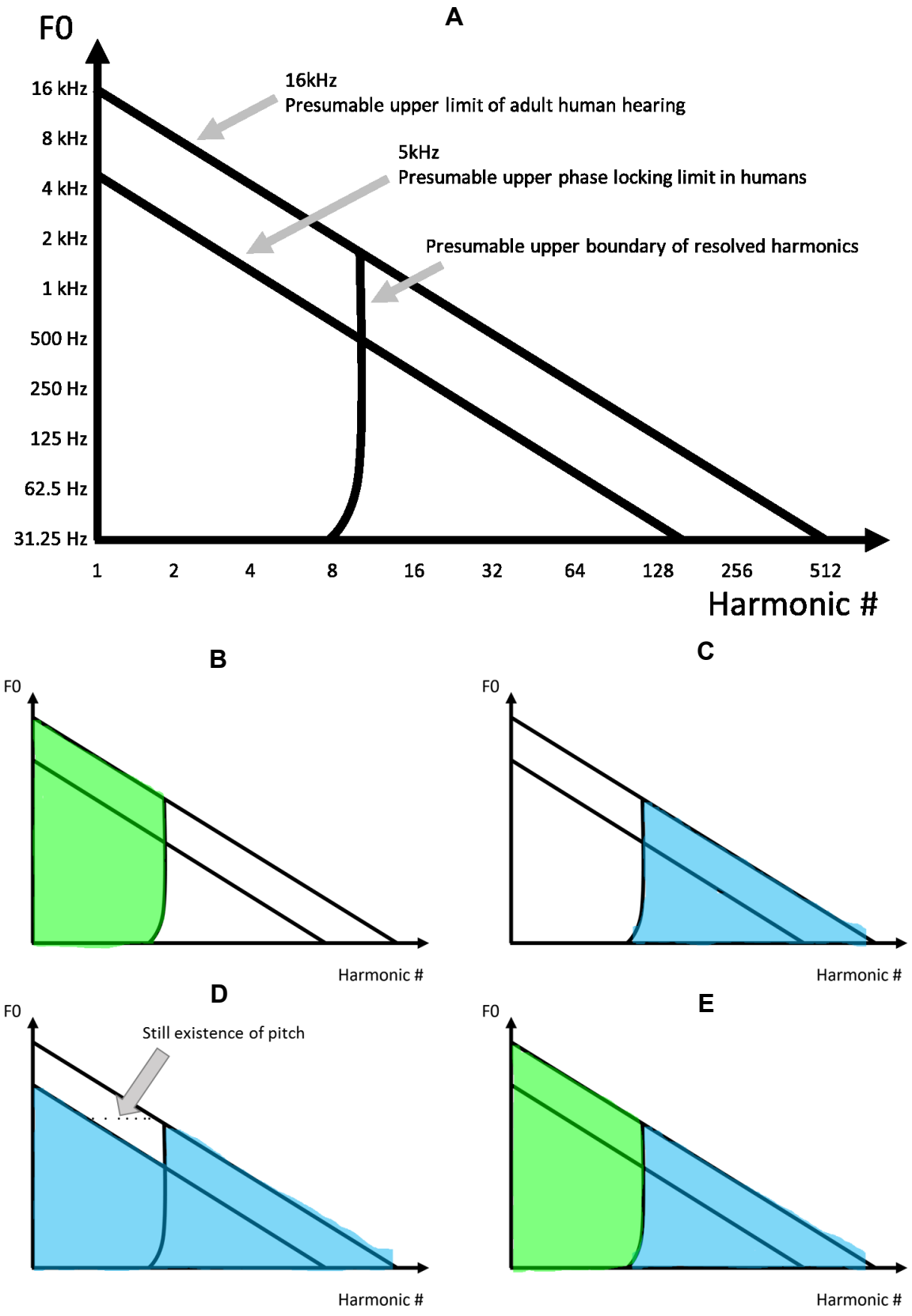

A
F0
16 kHz
8 kHz
4 kHz
2 kHz
1 kHz
500 Hz
250 Hz
125 Hz
62.5 Hz
31.25 Hz
16kHz
Presumable upper limit of adult human hearing
5kHz
Presumable upper phase locking limit in humans
Presumable upper boundary of resolved harmonics
1
2
4
8
16
32
64
128
256
512
Harmonic #
B
F0
Harmonic #
C
F0
Harmonic #
D
F0
Still existence of pitch
Harmonic #
E
F0
Harmonic #

Figure 1.2. Theoretical existence regions of pitch related cues and candidate pitch perception mechanisms

(A) The existence region map with the presumable upper hearing limit (~16 kHz for adult humans), the presumable phase locking limit (~5 kHz), and the upper boundary of resolved harmonics. The Y-axis shows different F0s, whereas the X-axis shows harmonic numbers. (B) The potential existence region for spectral harmonic template based pitch theory [Goldstein 1973, Terhardt 1974, Shamma and Klein 2000]. (C) The potential existence region for temporal envelope pitch theory [Schouten 1938]. (D) The potential existence region for temporal autocorrelation pitch theory [Meddis and O'Mard 1997, 2006]. Experiments show that resolved harmonics with frequencies above presumable phase locking limit can still evoke a robust pitch perception [Oxenham et al 2011], which is outside this existence region. (E) The potential existence region of a dual-mechanism pitch theory.

2. MATERIAL AND METHODS

2.1. Summary

In this chapter, the basic designs of behavioral task and operant conditioning methods are introduced. Detailed sound stimulus designs are more closely related to each specific experiment and thus are discussed separately with each experiment in chapter 3.3 and chapter 4.3.

And before designing any sound, two additional questions are discussed in this chapter as well. The first one is estimating the sound level of the distortion product at F0 when a missing fundamental complex sound is played to a marmoset. The second one is for any given harmonic, whether it is considered as a resolved harmonic or unresolved harmonic. These two questions are discussed in chapter 2.4 and 2.5. These two estimations are important to later chapters' experimental design.

2.2. General Experimental Methods

Details of marmoset operant conditioning tasks and apparatus can be found in recent publications from our laboratory [Osmanski and Wang, 2011, Remington et al. 2012]. All experimental procedures were approved by the Johns Hopkins University Animal Care and Use Committee and were in compliance with the guidelines of the National Institutes of Health.

2.2.1 Subjects

The subjects used in the experiments of this dissertation were adult marmosets ranging from 2 to 6 years old during testing. Most subjects had at least eight months experience in discrimination tasks, either with auditory peripheral tuning bandwidths measurements [Osmanski et al. 2013] or with pure tone discrimination training.

Subjects were housed in individual cages in a large colony at the Johns Hopkins University School of Medicine. All subjects were maintained at approximately 90% of their free-feeding weight on a diet consisting of monkey chow, fruit, and yogurt and had *ad libitum* access to water. Subjects were tested five or six days per week between the hours of 0900 and 1800. Each experimental session lasted approximately 60 minutes but no more than 100 minutes.

2.2.2 Apparatus

During testing, marmosets were seated in a custom Plexiglass restraint chair mounted in the center of a single-walled sound isolation chamber (Industrial Acoustic Company, Model 400A [101 cm (W) × 183 cm (D) × 198 cm (H) cm interior dimensions]) with the inside wall of the chamber lined with 7.5 cm thick acoustic absorption foam (Pinta Acoustics, model PROSPEC). Sound stimuli were played through a loudspeaker (Tannoy, model Arena, powered by amplifier Crown, D-75A), mounted 40~50 cm away in front of the animal's head.

All acoustic stimuli were generated offline using Matlab software (Mathworks, Natick, MA) and delivered at a nominal sampling rate of 100 kHz through a multi-processer DSP unit (Tucker-Davis Technologies, Alachua, FL, RX6), followed by a

programmable attenuator (Tucker-Davis Technologies, PA5), and an audio amplifier (Crown Audio, Model D-75A).

Liquid reward (a mixture of Gerber single-grain rice cereal, strawberry and/or banana-flavored Nesquik, and a protein powder supplement) was delivered through a food delivery tube attached to the top of the restraint chair and connected to a syringe pump (New Era Pump Systems, Inc., Farmingdale, NY, model NE-500) mounted to the base of the chair. Subject responses were recorded by monitoring when an infrared photobeam, positioned on a custom bracket at the end of the feeding tube, was broken by the subject licking at the feeding tube. Testing sessions were computer-controlled and monitored via webcam video (Logitech, C905 camera).

2.2.3 Stimulus calibration

Stimuli were calibrated using a 1/2-inch free-field microphone (Brüel & Kjaer, Type 4191) positioned in the chamber at the same location as the animal's head. The output of the microphone was amplified using a custom preamplifier, sent directly into a digital signal processor (Tucker-Davis Technologies, RX6), and analyzed using a custom Matlab calibration program written specifically for this hardware configuration.

2.3. Behavioral Task

In a basic behavioral task, animals were presented with repeating "background" sounds that had a fixed F0 or "equivalent" F0 (defined as the average frequency spacing between adjacent harmonics). Each testing trial had a variable duration waiting time that lasted from 3 to 15 seconds, during which the background sound was repeatedly

presented to the animal. After this waiting period, a “target” sound, which was always higher in F0 than the background, began to alternate with the background sound. Both the background and target sounds had a duration of 200ms with 10ms linear ramps (rise/fall time). The inter-stimulus interval during the task was fixed at 300ms. Animals could respond any time during the alternation period (i.e., the response window), which lasted for 4.8 seconds in total. The subject had to detect the “F0” change and respond by licking at a feeding tube placed in front of its mouth during the response window (“hit”) to receive food reward. However, if the subject licked before the response window onset, the chamber light was extinguished for 2-5 seconds as a warning signal, and the trial was restarted. If the subject did not respond during the trial at all, a “miss” was recorded and the system automatically started the next trial. This basic behavioral paradigm is illustrated in figure 2.1. Each experimental session generally contained at least 100 but not more than 200 trials.

To ensure that all subjects were attending primarily to F0 change to correctly perform the task, the sound level of each background sound presentation was randomized within a ±3 dB range. The level of each target sound (which alternated with the reference sounds) was always fixed.

2.3.1 Discrimination limen measuring task and analysis

The task paradigm is illustrated in figure 2.1.

In a discrimination limen measuring task, 70% of trials measured hit rates to real targets randomly chosen from 7 possible target choices, corresponding to seven different equivalent F0 distances from the background sound. These possible F0 changes were

equally spaced on the semitone scale (1 octave = 12 semitones, one semitone equals ~6% increase in periodicity), and chosen to bracket the presumed threshold (example see in figure 2.2). The remaining 30% of trials were sham trials in which no "target" sound was presented. Sham trials were used to measure false alarm rate as an indicator of how much the subject relied on guessing during the task. Sample raw hit rates are shown in figure 2.2 (A) for subject M13W under an unresolved harmonics condition (specific stimuli design will be discussed in chapter 4). As the F0 difference decreases, the raw hit rate drops from nearly perfect 100% to around the false alarm rate.

There are generally two ways to define discrimination thresholds, based on either corrected hit rate [Geschieder 1985], or d' - the signal detection sensitivity index [Green and Swets 1966].

The corrected hit rates were calculated from the raw values based on the false alarm rate according to the following equation: *Corrected hit rate = (Raw hit rate – False alarm rate) / (1 - False alarm rate)*. Discrimination thresholds were defined as that "F0" difference that the animals correctly identified 50% of the time (using linear interpolation of the corrected hit rate in each condition, figure 2.2 (A)) [Geschieder 1985].

Alternatively, signal detection sensitivity index (d') was estimated based on several assumptions: (1) the signal distributions of both background and target are normally distributed; (2) the signal distributions of both background and target have the same variance. Then d' was calculated by *d' = norminv(Raw hit rate) - norminv(False alarm rate)* in Matlab (Mathworks, MA, R2016a). Discrimination thresholds were defined as that "F0" difference where the animal produces a d'=1 (using linear

interpolation or extrapolation of the d' in each condition, figure 2.2 (B)) [Green and Swets 1966].

Thresholds were measured on the semitone scale under either pure tone conditions as frequency discrimination limens (FDL), or complex sound conditions as fundamental frequency discrimination limens (F0DL). Absolute FDL or F0DL in Hz were converted by the following equation: $Threshold_{in\ Hz} = Reference\ frequency \cdot (2\ ^{\wedge}\ (Threshold_{in\ semitone}/12)-1)$, where reference frequency is the periodicity or absolute frequency of the background sound. Relative thresholds in percentage were calculated by $Threshold_{in\ \%} = (Threshold_{in\ Hz} / Reference\ frequency) \cdot 100\%$.

Experimental sessions with a false alarm rate >25% or with a corrected hit rate curve passing below 50% multiple times were excluded from analyses. Testing continued until at least three consecutive sessions produced discrimination thresholds (corrected hit rate based) within one F0 spacing between adjacent targets.

2.3.2 Generalization task

A generalization task was modified from a discrimination limen measuring task (illustrated in figure 2.3). During the initial training phase of a generalization task, instead of seven different target sounds corresponding to seven different equivalent F0 distances from the background sound used in a discrimination limen measuring task, a single target sound is used. Furthermore, one out of every seven "target" trials with the target sound playing was a "probe" trial which provided no reward, even when subjects successfully detected the target sound from the background sound. Similarly, one out of the three "sham" trials with only background sound playing was another type of "probe" trial that

had no light-off given when subjects responded within the “sham” response window. Both types of “probe” trials are shown as the middle blocks in figure 2.3. Together, a task had 20% of total trials as real “sham” trials to measure false alarm rate, 60% of total trials as real “target” trials to measure hit rate, 10% of total trials as “probe 1” trials just as “sham” trials but without reward or light-off, 10% of total trials as “probe 2” trials just as “target” trials but without reward or light-off.

After the initial training procedure, which familiarizes subjects with the probability that around six out of seven target trials come with reward, the task was switched into the testing procedure (paradigm shown in figure 2.4). The “sham target” sound (the same as “background” sound) used inside the “probe 1” trials’ response window was altered into a “probe 1” sound that was different form both “background” sound and “target” sound. The “target” sound used inside the “probe 2” trials’ response window was altered into a “probe 2” sound that was also different from both “background” sound and “target” sound. If an altered “probe x” sound was perceptually close to the “background” sound, then the subject’s hit rate on that type of “probe” trials should be similar to the false alarm rate. If an altered “probe x” sound was perceptually close to the “target” sound, then the subject would not be able to differentiate this type of “probe” trial from real “target” trials, and thus the hit rate for these “probe” trials would be comparable to that of real “target” trials. This generalization task paradigm is illustrated in figure 2.4.

To quantify the perceptual distance of a sound from the background, d’ was calculated by *d’ = norminv(Raw hit rate) - norminv(False alarm rate)* in Matlab

(R2016a, Mathworks, MA). The calculation was repeated for target trials, probe 1 trials, and probe 2 trials.

2.4. Estimation of Nonlinear Distortion Product

As introduced in the chapter 1.2.1, due to the cochlear nonlinearity, a missing fundamental sound may reintroduce a F0 component on the basilar membrane in the cochlea, through nonlinear interactions between adjacent harmonics. The question is, how loud is this distortion product?

Assuming two adjacent harmonics are f_1, and f_2. The reintroduced frequency component by quadratic nonlinear distortion is at a frequency of f_2-f_1. And the level of this nonlinear distortion product generally follows $L(f_1) + L(f_2) – a(f_1, f_2)$, where $L(f_1)$ and $L(f_2)$ are sound level of f_1 and f_2 frequencies, and $a(f_1, f_2)$ is a constant (in dB) when f_1, f_2, and the relative phase between them are fixed.

The level of this quadratic distortion tone (QDT) is hard to measure directly in humans. However, through psychoacoustical procedures, a subject can report whether the QDT is perceptually prominent or not. Assuming the QDT is at a certain phase at the frequency of f_2-f_1, if a tone of the same frequency but in the opposite phase to the QDT is delivered simultaneously to the subject with f_1 and f_2, the QDT can be perceptually diminished or cancelled. Studies have shown that the level of the QDT can be inferred by the level of such a cancellation tone when subjects are asked to adjust both the frequency and phase of this tone to maximize the cancellation effect [Goldstein 1967, Oxenham AJ et al 2009, Pressnitzer and Patterson 2001]. Based on the inferred QDT level, the constant $a(f_1, f_2)$ is estimated as $L(f_1) + L(f_2) – L(QDT)$, as shown in figure 2.5. The mean value of

this constant is between 105 dB and 110 dB, with maximal and minimal values as 122 dB and 98 dB respectively. This suggests that in humans if f_1, f_2 are played at both at 50 dB SL, then the QDT is very likely lower than the hearing threshold (0 dB SL), and is hardly to be heard.

In nonhuman animals, although a similar psychoacoustic approach may not be as feasible as in humans, invasive physiological methods are more accessible. To estimate the level of QDT, auditory single units were recorded along the auditory pathway, including the anteroventral cochlear nucleus (AVCN) [Smoorenberg et al 1976, Faulstich and Kossl 1999], and inferior colliculus (IC) [McAlpine 2004, Abel and Kossl 2009]. For a certain auditory unit, it's receptive field properties were determined. Then, a complex tone outside the neuron's receptive field was designed and delivered, presumably to generate a QDT that can appear back inside its receptive field. The tuning properties of the complex tone were compared to the original receptive field properties, and the level of $a(f_1, f_2)$ can be estimated based on these property comparisons. The estimated levels from cats, gerbils, and guinea pigs were shown in figure 2.6. Most studies have the constant $a(f_1, f_2)$ higher than 60 dB. In other words, if f_1, f_2 are both played at 30 dB SL, QDT is hardly detectable neurophysiologically in these studies.

Due to the existence of the QDT, a doubt was casted [McAlpine 2004] on several previous proposals of pitch-like neurons, recorded in animal IC or auditory cortex [Langner G 1997, Schulze H and Langner G 1997, Schulze H and Langner G 1999, Biebel and Langner 2002, Schulze et al 2002], since none of these studies showed evidence to exclude the possibility that the tuning and the sensitivity to missing fundamental sounds may be due to QDTs.

To sum up, if two adjacent harmonics were played each at 50 dB SL, the evidence from previous human psychoacoustic studies shows the level of QDT is likely below the psychophysical detection threshold (0 dB SL) for humans. And if two adjacent harmonics were played each at 30 dB SL, the evidence from previous nonhuman mammal studies shows that the QDT becomes hardly detectable through neural recordings, in species such as cats, gerbils, and guinea pigs. Those numbers provide a guideline for the stimuli design in the following two chapters.

2.5. Estimation of Harmonic Resolvability

Before starting to investigate into the mechanisms of complex sound pitch perception in a new species, one first needs to know which harmonic is the highest resolvable harmonic for any given f0. As introduced in the chapter 1.3, the existence region of a candidate pitch theory is highly dependent on peripheral harmonic resolvability. If the tuning bandwidths of a species' auditory filters are constantly sharp across the hearing range, then all harmonics would be resolved and there would be little room left for a temporal envelope based theory. On the other hand, if the tuning bandwidths of a species' auditory filters are constantly wide across the hearing range, then every harmonic is hardly separable from its adjacent ones and a spectral template based mechanism is unlikely to exist. In the following part, I will first compare frequency resolution in different mammal species, and then discuss how to estimate peripheral harmonic resolvability boundaries based on peripheral frequency resolution for a given f0 in a certain species.

2.5.1 The comparative analysis of frequency resolution of auditory peripheries

The frequency resolution at the auditory periphery can be quantified as tuning bandwidths of auditory filters. In practice, this resolution can be quantified as the equivalent rectangular bandwidth (ERB) [Moore and Glasberg 1990] at different frequencies through several different methods. A classical psychoacoustic approach has estimated tuning bandwidths of auditory filters in humans [Moore and Glasberg 1990], marmosets [Osmanski et al 2013], ferrets [Alves-Pinto et al 2016], chinchillas [Niemiec et al 1992], guinea pigs [Evans et al 1992], and mice [May et al 2006] by measuring the detectability of a tone delivered inside a spectrally notched noise (data shown in figure 2.7). Neurophysiological recordings of auditory nerve have also shown tuning bandwidths in macaques [Joris et al 2011], cats [Joris et al 2011, Shera et al 2010, Cedolin and Delgutte 2005], guinea pigs [Tsuji and Liberman 1997], and chinchillas [Recio-Spinoso et al 2005] (data shown in figure 2.8). More recently, a noninvasive physiological method utilizing stimulus frequency otoacoustical emission (SFOAE) was also developed to estimate tuning bandwidths in humans [Shera et al 2002], macaques [Joris et al 2011], marmosets [Bergevin et al 2011], cats [Shera and Guinan 2003], guinea pigs [Shera and Guinan 2003], and chinchillas [Siegel et al 2005] (data shown in figure 2.9).

Across the data from these three different methods, it is generally consistent that the absolute frequency tuning bandwidth of auditory peripheries follow: humans < monkeys < carnivores < rodents and the tuning sharpness follows: humans > monkeys > carnivores > rodents.

2.5.2 The convergence of the definition of resolvability

As introduced above, the absolute frequency resolution at a particular frequency at the auditory periphery can be quantified by the tuning bandwidth of an auditory filter centered at that frequency, measured as the equivalent rectangular bandwidth (ERB). The relative resolution for a specified F0 can be described as the ratio α between this F0 and the measured ERB at the frequency, as $\alpha = F0/ERB$, or $ERB = 1/\alpha \cdot F0$. The smaller α goes, the more harmonics can pass through this auditory filter's bandwidth, thus increasing the likelihood that adjacent harmonics become unresolved. Alternatively, the higher α goes, fewer harmonics (or none) pass within the bandwidth of the auditory filter, thus increasing the likelihood that adjacent harmonics become resolved.

Resolved harmonics, by definition, can be heard out individually from the complex. In humans, the highest resolved harmonic of a wide range of F0s was assessed behaviorally [Plomp 1964, Plomp and Mimpen 1968]. These data show that, for a particular F0, it's proportional to the ERB at the frequency of the highest resolved harmonic, with a fixed ratio α around 1~1.25 [Moore 2012]. This ratio provides a link between peripheral frequency resolution and the upper boundary of resolved harmonics.

Another path for deriving the link between peripheral frequency resolution and harmonic resolvability boundaries is to examine the saliency of temporal envelope cues. When adjacent harmonics have an alternating phase relationship (e.g., sine and cosine starting phase), the amplitude profile of the sound spectrum remains unchanged but the overall periodicity is doubled compared to the case when all harmonics begin in the same phase. If a sound's temporal envelope is the dominant cue for perceiving pitch, then an alternating phase complex should be reported as bearing a pitch an octave higher than a

purely sine/cosine phase complex. Indeed, this was observed on unresolved harmonics but not on resolved harmonics [Shackleton and Carlyon 1994]. The resolvability of a harmonic was defined in terms of the number of harmonics contained between the 10 dB down points (~1.8 ERB) of an auditory filter centered on that harmonic. If the number is lower than 2, it's resolved. If the number is higher than 3.25, it's unresolved. The upper boundary of resolved harmonics and the lower boundary of unresolved harmonics were thus estimated having α values as 0.9 and 0.55 respectively.

In addition, a recent study estimated the upper boundary of resolved harmonic to be where the peak to valley ratio of the excitation pattern is 1.98 dB [Bernstein and Oxenham 2006], from which α = 1.18 can be derived mathematically.

Based on the findings described above, I propose α values of 1 and 0.55 as the upper boundary of resolved harmonics and lower boundary of unresolved harmonics, respectively.

2.5.3 Resolvability boundaries in marmosets

The absolute frequency resolution in marmosets was measured as ERB using the same psychoacoustical approach as used in humans [Osmanski et al 2013].

A one-parameter rounded exponential filter was borrowed from the human excitation pattern model [Moore and Glasberg 1983] to build an analogous model of the marmoset excitation pattern. ERBs were interpolated ('pchirp' in Matlab R2013b, Mathworks) from behavioral measurements of peripheral tuning bandwidths in marmosets [Osmanski et al 2013] to cover the entire marmoset hearing range. Parameter p in the rounded exponential filter model is given as $p = 4 \cdot f / ERB$ [Moore and Glasberg

1983]. A harmonic complex tone of a F0 of 440 Hz and composed by up to the 64th harmonic was drawn in figure 2.10 (A) as the vertical lines. The shadowed area indicates its excitation pattern at marmoset auditory peripheries.

The spatiotemporal activity pattern (cochleogram) was also modeled to show both resolvability change across different frequencies and temporal envelope cue along the temporal axis, as shown in figure 2.10 (B). Auditory filters were modeled based on the Gammatone filter bank algorithm [Slaney 1998]. Bandwidth parameters were also taken from marmoset ERB data [Osmanski et al 2013]. The incoming sound signal was passed through this filter bank and then rectified. Vertically oscillating stripes are discretely distributed on each resolved harmonic. However, on higher frequency side, a harmonic is no longer spectrally separable from adjacent harmonics. In return, their temporal interactions generate a temporal envelope repetition rate equal to F0, and can serve as a pitch cue.

To derive resolvability boundaries based on these ERB data for marmosets. α values of 1 and 0.55 as discussed above were used. The resolvability boundaries in humans were plotted as the grey dashed lines in figure 2.11 (A), whereas black solid lines indicate the same boundaries estimated in marmosets. The α values of 1 for the upper boundary of resolved harmonics was validated by tuning bandwidths of single units recorded at each unit's best sound level in awake marmoset auditory cortex [Bartlett et al 2011, Osmanski et al 2013]. For an F0 equal to 440 Hz, marmoset shows resolvability comparable to humans. Figure 2.11(B) shows the excitation patterns of resolved and unresolved harmonics, in which fluctuations can be seen on resolved harmonics (green) but not on unresolved harmonics (blue).

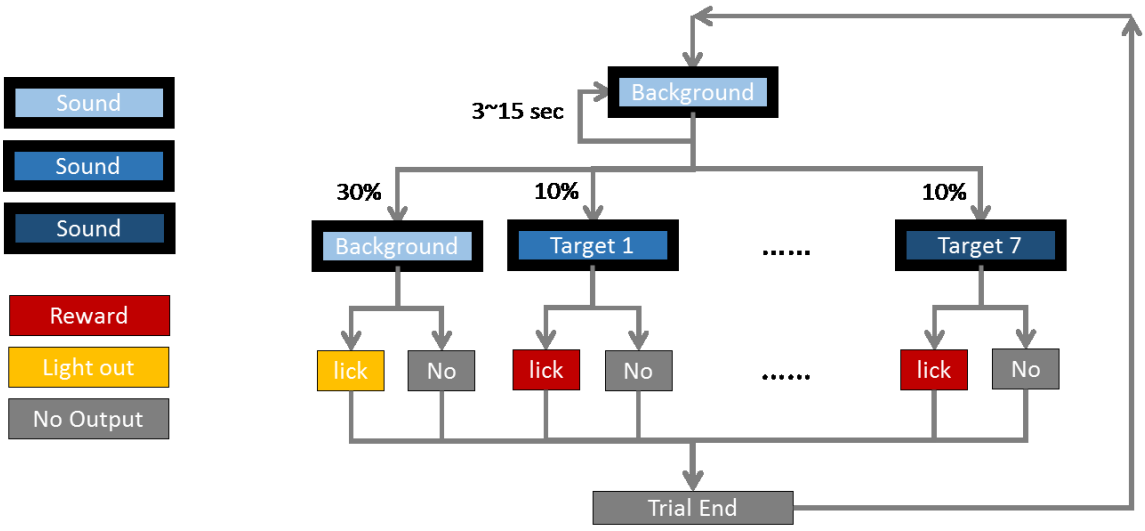


Figure 2.1 Behavior paradigm of a discrimination limen measuring task

Each testing trial had a variable waiting period that lasted between 3 to 15 seconds, during which the background sound was repeatedly presented to the subject. After this waiting period, a “target” sound, which was always higher in F0 than the background sound, began to alternate with the background sound. The subject could respond any time during the alternation period (i.e., the response window), which lasted for 4.8 seconds in total. The subject had to detect the “F0” change and respond by licking at a feeding tube placed in front of its mouth during the response window (“hit”) to receive food reward.

In a discrimination limen measuring task, 70% trials were measuring hit rates on real targets randomly chosen from 7 possible target choices, corresponding to seven different F0 distances from the background sound. The remaining 30% of trials were sham trials in which no "target" sound was presented. Sham trials were used to measure

false alarm rate as an indicator of how much the subject relied on guessing during the task.

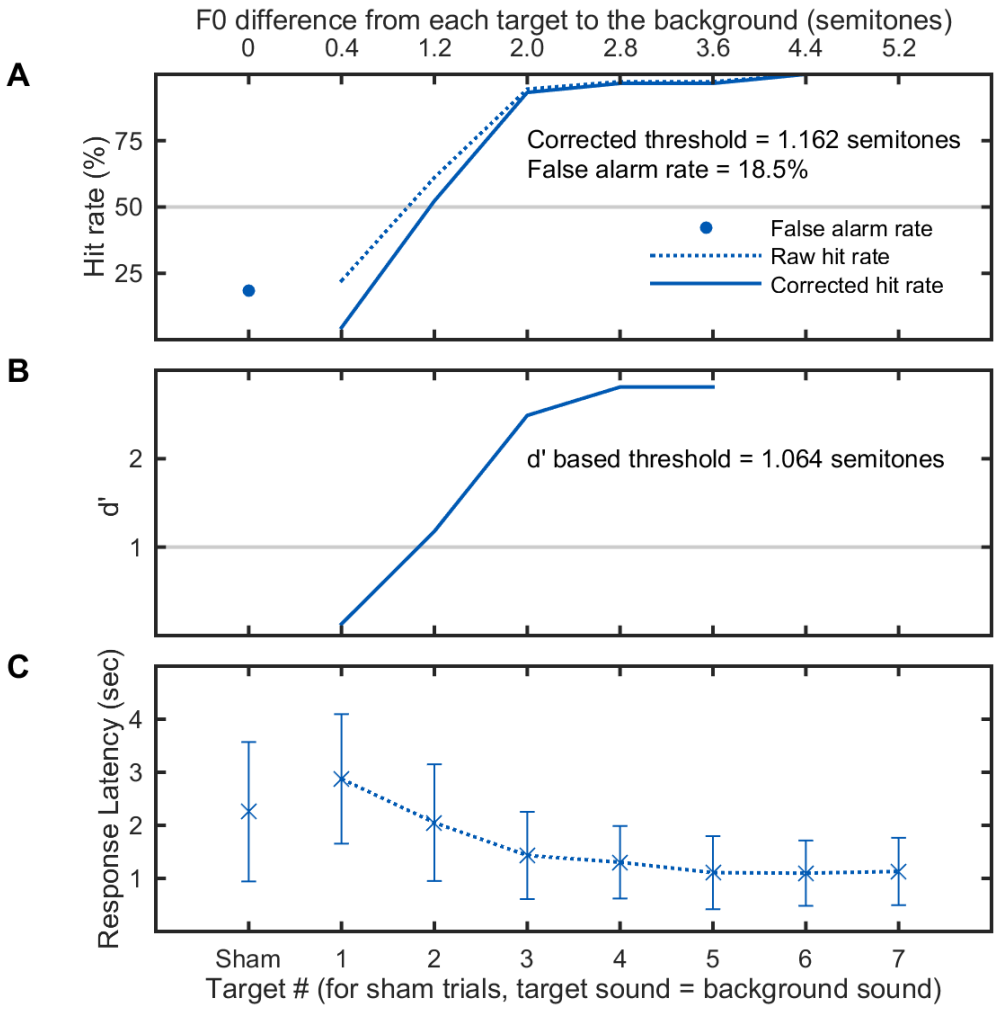


Figure 2.2 Summary of discrimination threshold calculations

(A) The corrected psychometric curve along 7 different targets in the 2nd F0DL measure, on subject M13W, under unresolved harmonics condition (see stimuli design in chapter 4.4.1). Raw hit rates and the false alarm rate are also shown. (B) The d' based psychometric curve along 7 different targets of the same measure. (C) Response latencies across different targets. False alarm response latency is also shown, labeled by "Sham".

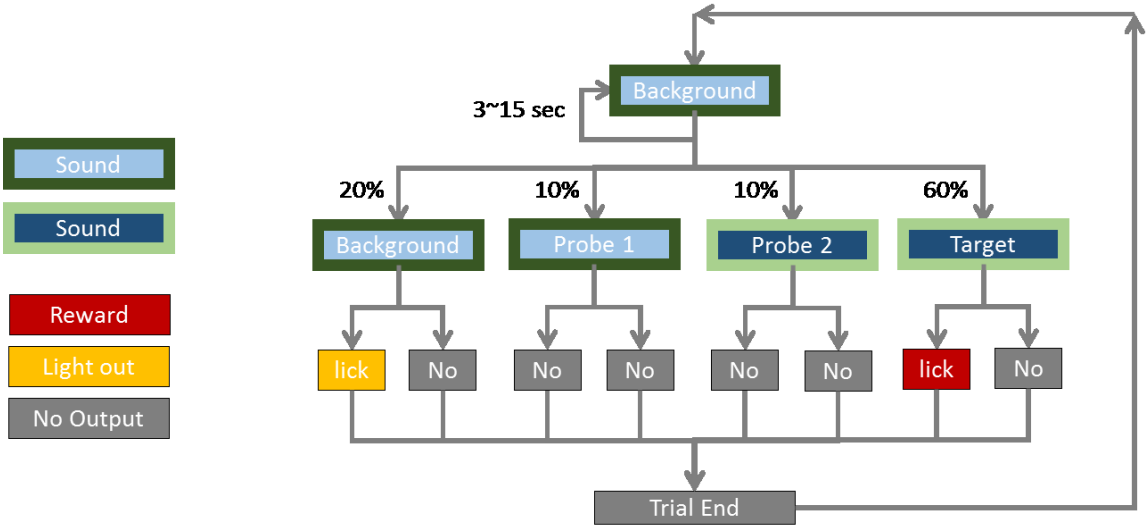


Figure 2.3 Behavior paradigm of a generalization task (training procedure)

Each trial had a variable waiting period that lasted between 3 to 15 seconds, during which this background sound was repeatedly presented to the animal.

In the training procedure of a generalization task, 20% of total trials are real "sham" trials to measure false alarm rate, 60% of total trials are real "target" trials to measure hit rate, 10% of total trials are "probe 1" trials just as "sham" trials but without reward or light-off, 10% of total trials are "probe 2" trials just as "target" trials but without reward or light-off.

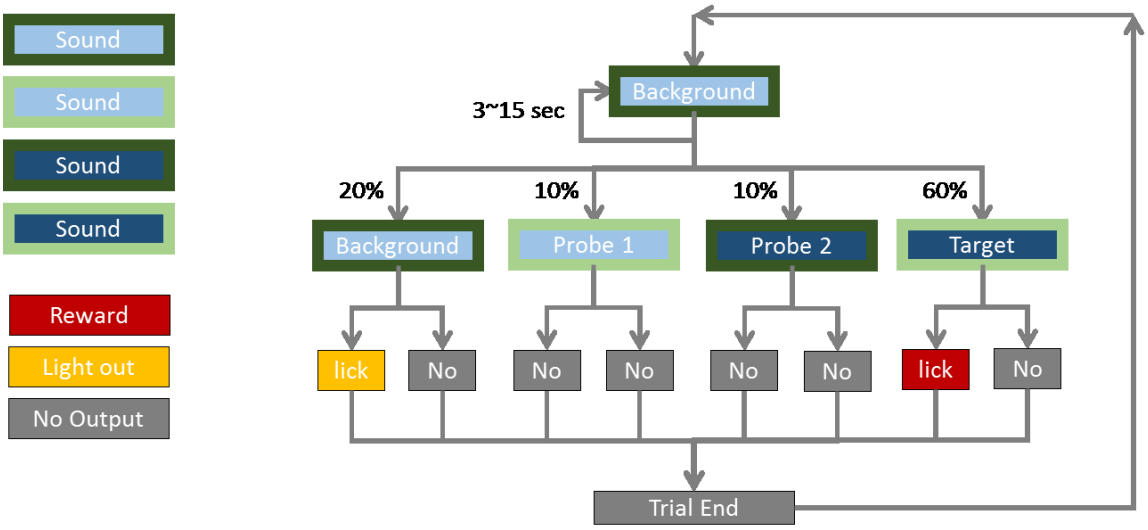


Figure 2.4 Behavior paradigm of a generalization task (testing procedure)

Each trial had a variable waiting period that lasted from 3 to 15 seconds, during which this background sound was repeatedly presented to the animal.

In the testing procedure of a generalization task, 20% of total trials are real "sham" trials to measure false alarm rate, 60% of total trials are real "target" trials to measure hit rate, 10% of total trials are "probe 1" trials without reward or light-off, 10% of total trials are "probe 2" trials without reward or light-off. Probe trials have testing probe sounds that are different from both background sound and target sound along certain acoustic dimensions.

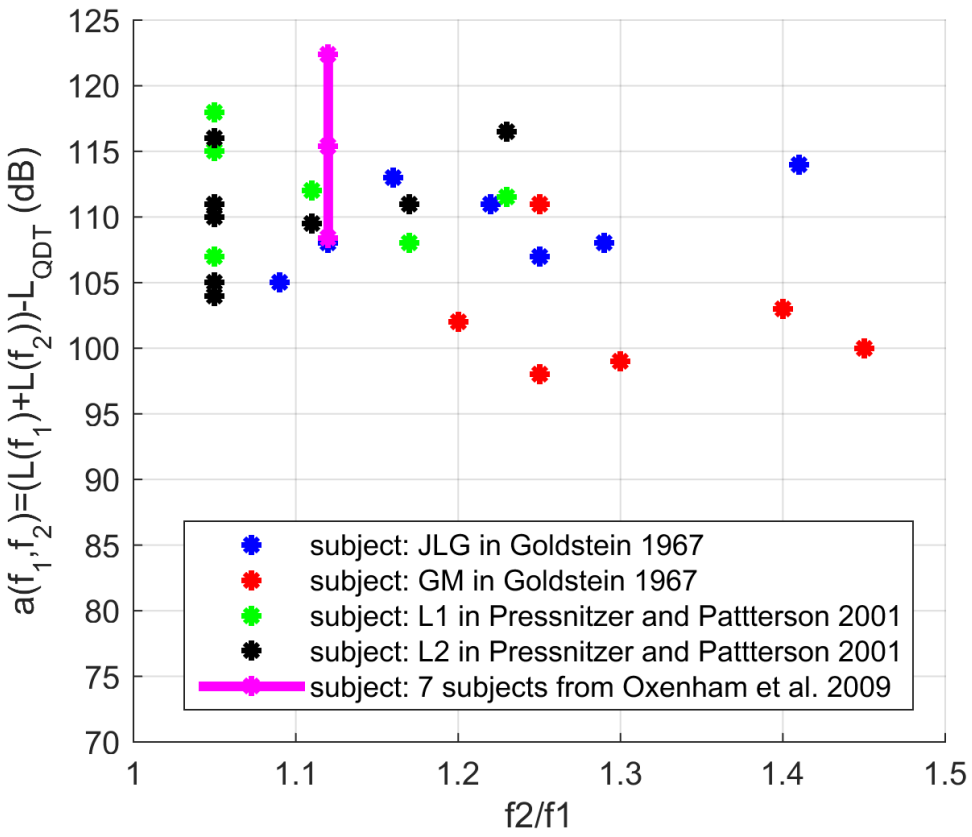


Figure 2.5 Summary of human cochlear distortion product estimations

The quadratic nonlinear distortion constant $a(f_1, f_2)$ was estimated in humans using psychoacoustic approaches and summarized in the plot. [Goldstein 1967, Oxenham AJ et al 2009, Pressnitzer and Patterson 2001].

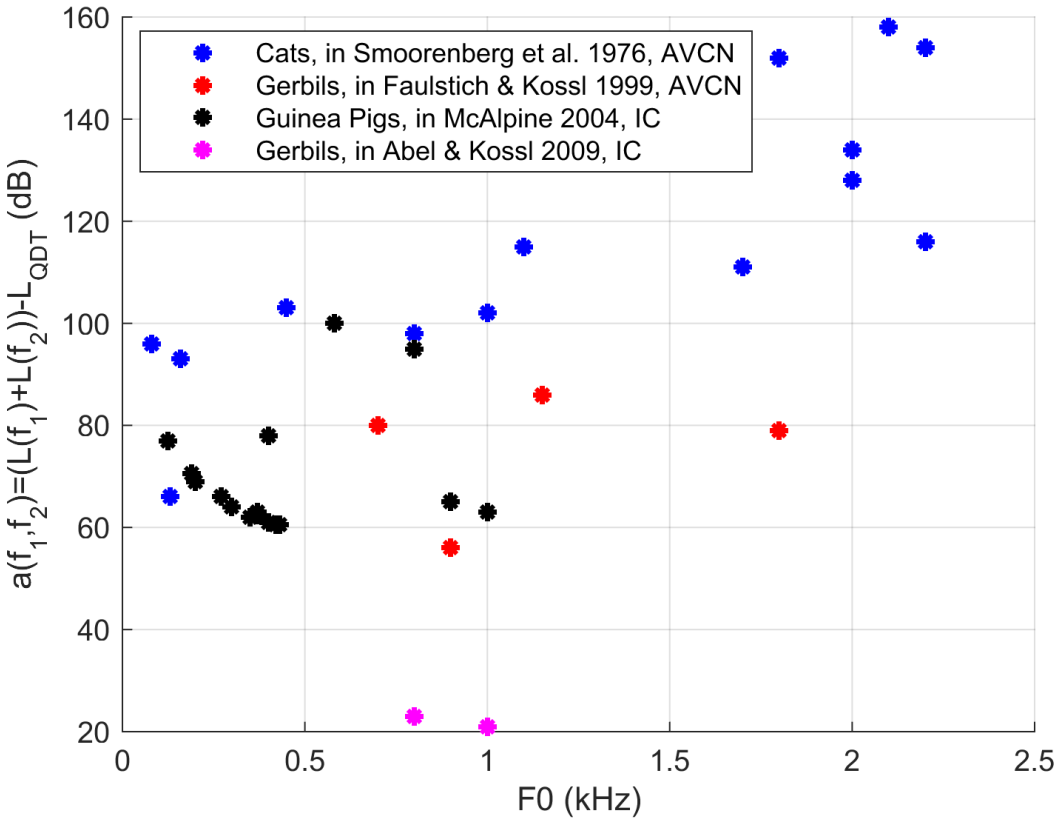


Figure 2.6 Summary of animal cochlear distortion product estimations

The quadratic nonlinear distortion constant $a(f_1, f_2)$ was estimated in nonhuman animals using neurophysiological approaches and summarized in the plot. [Smoorenberg et al 1976, Faulstich and Kossl 1999, McAlpine 2004, Abel and Kossl 2009].

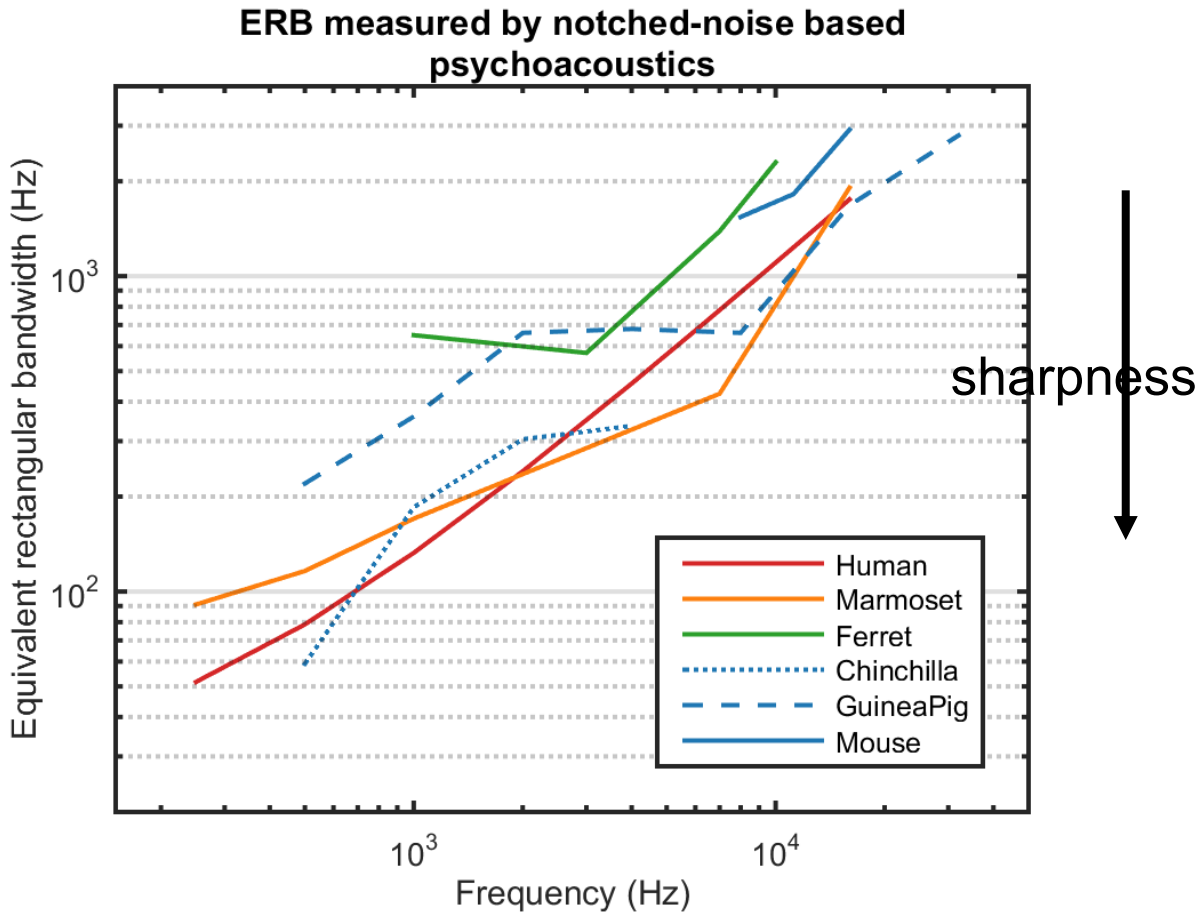


Figure 2.7 Auditory filters' sharpness comparison, behavioral measures from notched-noise psychoacoustic experiments

Tuning bandwidths of auditory peripheries are shown in equivalent rectangular bandwidth (ERB) measured from humans [Moore and Glasberg 1990], marmosets [Osmanski et al 2013], ferret [Alves-Pinto et al 2016], chinchillas [Niemiec et al 1992], guinea pigs [Evans et al 1992], and mice [May et al 2006]. As the smaller ERB is, the sharper the auditory filter is.

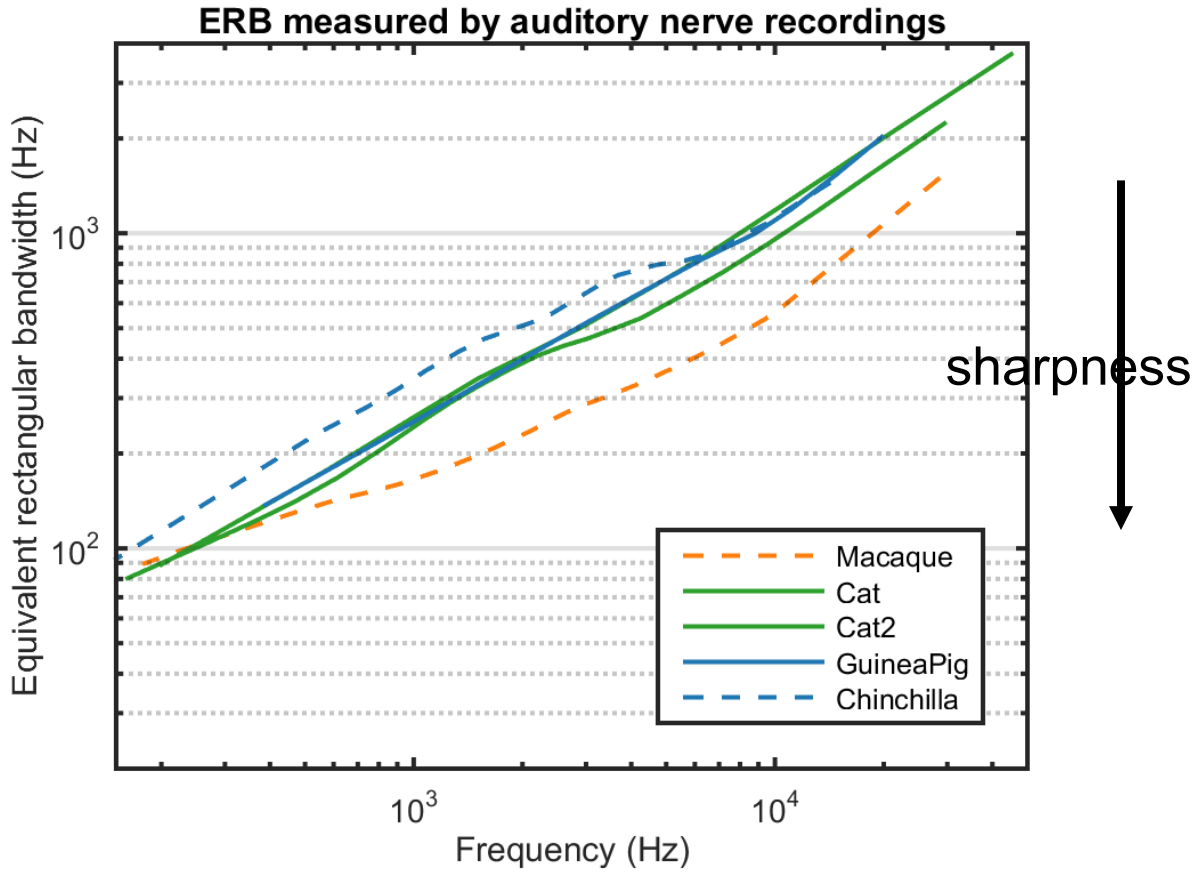


Figure 2.8 Auditory filters' sharpness comparison, physiological measures from auditory nerve recordings

Tuning bandwidths of auditory peripheries are shown in equivalent rectangular bandwidth (ERB) measured neurophysiologically from auditory nerves in macaques [Joris et al 2011], cats [Joris et al 2011, Shera et al 2010, Cedolin and Delgutte 2005], guinea pigs [Tsuji and Liberman 1997], and chinchillas [Recio-Spinoso et al 2005]. As the smaller ERB is, the sharper the auditory filter is.

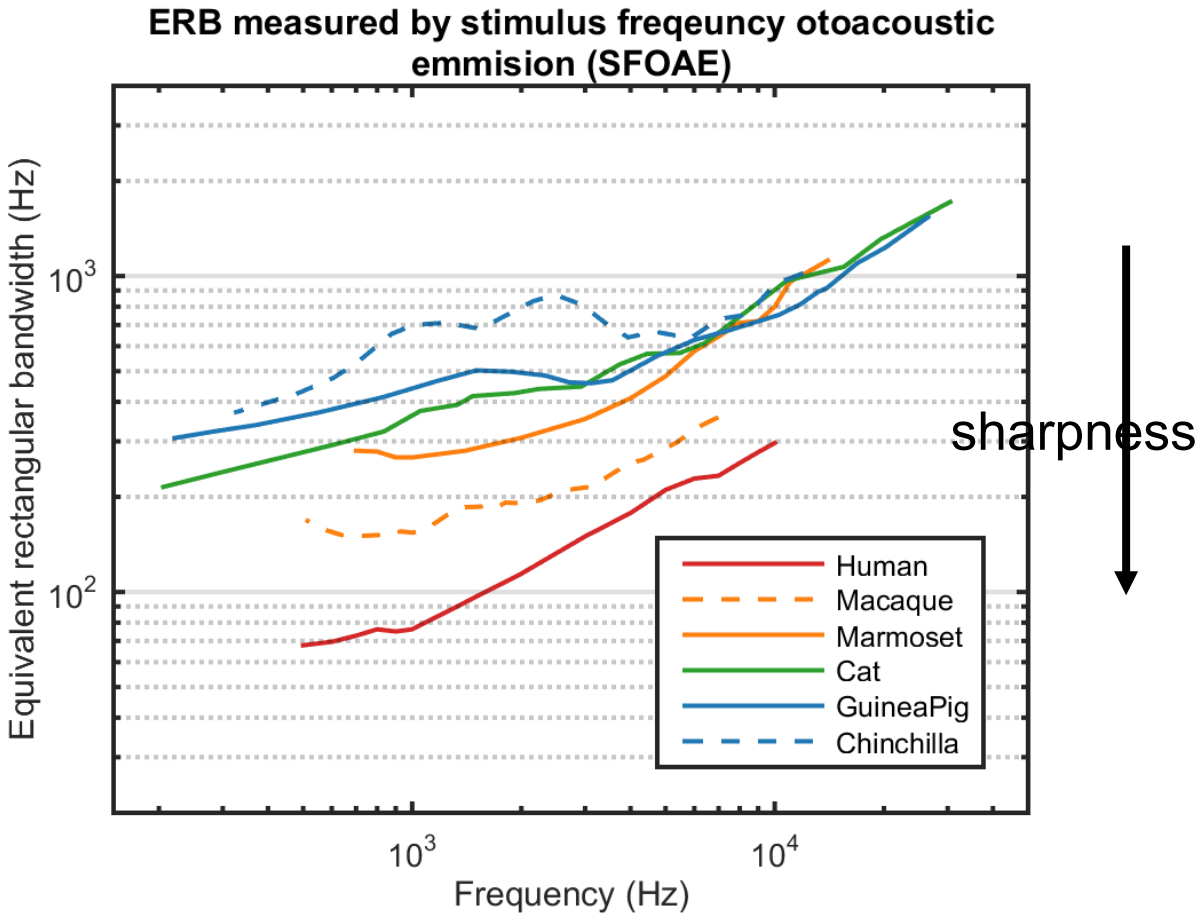


Figure 2.9 Auditory filters' sharpness comparison, physiological measures from stimulus frequency otoacoustic emission (SFOAE) experiments

Tuning bandwidths of auditory peripheries are shown in equivalent rectangular bandwidth (ERB) estimated based on SFOAE measurements in humans [Shera et al 2002], macaques [Joris et al 2011], marmosets [Bergevin et al 2011], cats [Shera and Guinan 2003], guinea pigs [Shera and Guinan 2003], and chinchillas [Siegel et al 2005]. As the smaller ERB is, the sharper the auditory filter is. The ratio Q_{ERB}/N_{BM} is assumed to be 1.25 [Shera et al 2010].

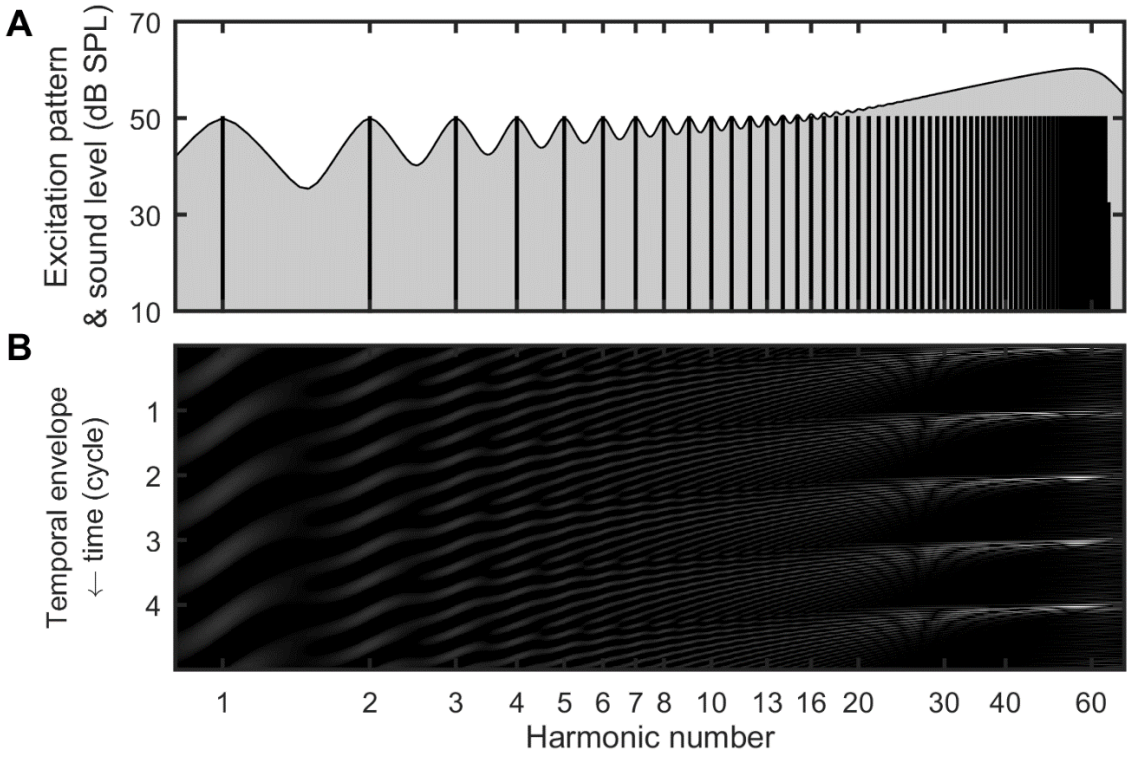


Figure 2.10 Harmonic resolvability in marmosets.

(A) Vertical lines indicate the acoustic spectrum and sound levels of the background sound used in "ALL" condition F0DL measurements (F0 = 440 Hz, up to the 64th harmonic). The shadowed area indicates its excitation pattern in marmoset auditory peripheries. (B) The spatiotemporal activity pattern of marmoset auditory peripheries. Five F0 cycles were shown along the vertical axis, against harmonic numbers along the horizontal axis. Vertically oscillating stripes are discretely distributed on each resolved harmonic. However, on higher frequency unresolved harmonic, harmonics can no longer be spectrally separable from adjacent harmonics. In return, their temporal interactions generate a temporal envelope repetition rate equal to F0 (horizontal stripes), and can serve as a pitch cue.

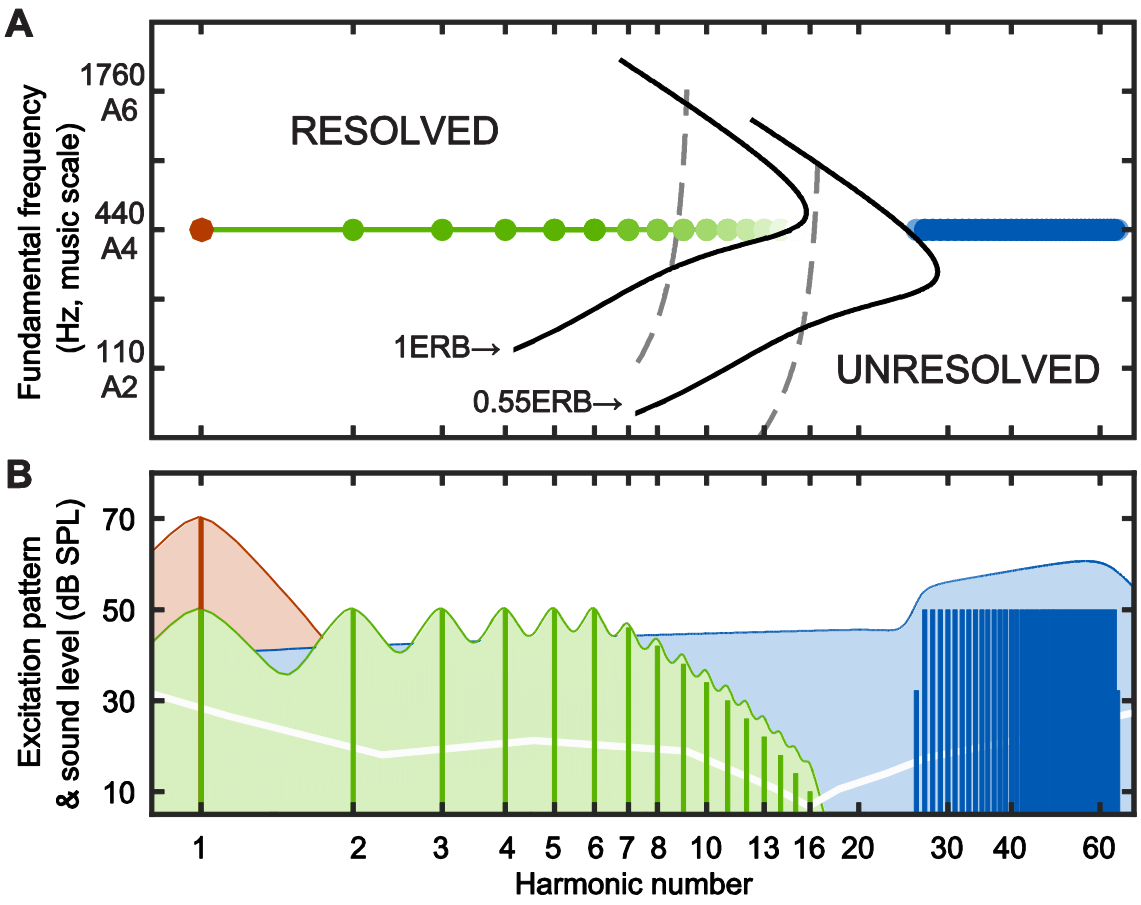


Figure 2.11 Summary of resolvability boundaries in marmosets.

(A) Estimated harmonic resolvability across F0s. Dashed grey lines indicate estimated upper boundary of resolved harmonics (1 ERB) and lower boundary of unresolved harmonics (0.55 ERB) in humans. Black lines indicate the same boundaries estimated in marmosets. At F0=440Hz, pure tone at F0 (red), resolved harmonics (green), unresolved harmonics (blue) are indicated by circles. (B) Illustration of excitation patterns of background sounds used in later F0DL measuring experiments. Vertical lines indicate acoustic spectra and sound levels of the background sounds used in marmoset

F0DL measurements. Colored areas indicate peripheral excitation patterns [Moore and Glasberg 1983] in marmosets, in which fluctuations can be seen on resolved harmonics (green) but not on unresolved harmonics (blue). Extended blue tail of unresolved harmonics on low frequency side indicates a noise masker used to mask potential distortion products from unresolved harmonics back into the resolved side when measuring unresolved harmonics' F0DL. The white line references marmoset audiogram [Osmanski and Wang 2011].

3. MISSING FUNDAMENTAL PERCEPTION

3.1. Summary

One of the most prominent features of human pitch perception is the missing fundamental phenomenon, which suggests the fundamental frequency (F0) component is not necessary for a salient pitch to be perceived from a complex sound. The pitch thus must be derived actively from non-fundamental harmonics by the auditory system. In humans, missing fundamental perception is precise enough for perceiving music melodies, which have smallest pitch step to be just one semitone.

In this chapter, marmosets were trained to robustly discriminate harmonic complex sounds that were different in both periodicity and the presence of an F0 component. When introducing a probing sound that was only different from the background sound in its periodicity, the hit rate remained high. When introducing another probing sound that was only different from the background sound in its F0 presence, the hit rate remained indistinguishable from the false alarm rate. These results suggested that marmosets discriminated these sounds based on their periodicity but not the presence of an F0 component, and thus possess missing fundamental pitch perception. All periodicity differences used in the current study were only one semitone (the smallest pitch step on the music scale). This is the first time that a nonhuman species has been shown to have missing fundamental pitch at this precision, and suggests that marmosets may potentially be able to discriminate musical melodies through this precise missing fundamental pitch perception. This human-like feature found in marmosets may guide further investigations of music element related perceptions in marmosets.

3.2. Introduction

Humans can perceive robust music melodies from missing fundamental complex sounds, even in some cases when all spectral components were restricted to those above 5 kHz [Oxenham et al 2011]. Sensitivity to the periodicity of missing fundamental sounds has been shown in several non-human species, including cats [Heffner and Whitfield 1976, Chung and Colavita 1976], songbirds [Cynx and Shapiro 1986, Okanoya 2000], macaques [Tomlinson and Schwarz 1988], goldfish [Fay 2005], and chinchillas [Shofner 2011]. Table 3.1 shows a summary of these demonstrations. However, none of these studies has demonstrated the perception is precise enough for one semitone difference in periodicity, which is the smallest pitch step on the music scale. Among these studies, the smallest pitch step used is three semitones. Here I want to ask whether marmoset monkeys also have missing fundamental pitch perception, and, if so, whether this perception is precise enough to potentially allow fine music melody discriminations.

3.3. Methods

The behavioral apparatus, generalization task, and related analysis methods have been discussed in chapter 2. In this part of the study, subjects were initially trained to detect the appearance of the target sound from the background sound. Both sounds were harmonic complex sounds. There were two major differences between the target sound and the background sound. The first difference was in their periodicity that the “target” sound had a periodicity that was just one semitone higher than the “background” sound, which had a periodicity equal to 440Hz, also known as A440 or A4 in musical tuning standard [ISO 16]. The periodicity of the target sound was around 466.16 Hz, as A#4 on

the music scale. The second difference was that the target sound was composed by harmonics but without the fundamental frequency (F0) component, whereas the background sound was composed by harmonics with the F0 component presented. Subjects could potentially learn the discrimination task based on the F0 difference, or the presence of the F0 component, or both.

To minimize the possibility that our subjects could use spectral envelope edges as a cue for discrimination, we implemented roll-offs on the spectral edges. Upper spectral edges of all stimuli sounds were rolled off starting from 50dB at 1320 Hz (third harmonic of 440 Hz F0) with a slow slope (4dB/ 440Hz), as $Level = (50 - (f - 3 \cdot 440Hz) \cdot 4 / 440Hz) \, dB \, SPL$, ending at 4840Hz. For all the stimuli, the maximum level of harmonics was calibrated to be around 50dB SPL, which is estimated to be around 30 dB SL based on the marmoset audiogram [Osmanski and Wang 2011]. The spectra used in the current studies are illustrated in figure 3.1.

To minimize the possibility that our subjects could use nonlinear distortion product at F0 as a cue to perceive the periodicity, we implemented a band-passed noise masker with cut-off frequencies at 100Hz and 800Hz, by passing white noise generated online through two second order Butterworth filters. The level of the noise masker was estimated at ~42 dB SPL / ERB at 440 Hz, based on the previously measured ERB data in marmosets [Osmanski et al 2013]. Two additional noise levels were tested, estimated to be ~34 dB SPL / ERB and ~22 dB SPL / ERB.

For training procedures [see chapter 2.3.2 and figure 2.3], “probe 1” trials have the same sound stimulus designs as the sounds used in real “sham” trials, whereas “probe 2” trials have the same sound stimulus designs as the sounds used in real “target” trials.

Subjects thus have no cue available to discriminate “probe 1” trials from “sham” trials and to discriminate “probe 2” trials from “target” trials, and learned to adapt to the probability of the reward presence. After subjects performed stably in the training procedures, they were switched to the testing procedures (see chapter 2.3.2 and figure 2.4).

For the testing procedures, “probe” trials’ sound stimuli were deviated from both the background sound and the target sound. The spectra of these sounds are shown in figure 3.1. “probe 1” trials had a probe sound which has the same periodicity (A4) with the background sound, but without the F0 component presented, like what in the target sound. “Probe 2” trials had another probe sound which has the F0 component presented, like the background sound, but with a periodicity (A#4) same as that of the target sound.

Hit rates of “probe 1” trials, “probe 2” trials, “target” trials were measured to estimate the perceptual distances from the sounds used in these trials to the background sound, using d’ calculation introduced in chapter 2.3.2.

3.4. Results

3.4.1 Existence and precision of missing fundamental pitch perception in marmosets

To test whether marmosets have missing fundamental perception, we first trained subjects to discriminate the target sound from the background sound with differences in two dimensions. The first is whether there is an increase in periodicity, the other is whether the F0 component is absent or presented. Subjects can potentially perform the discrimination task based on the difference along perceptual dimensions of periodicity change, or F0 presence, or both.

After subjects learned to discriminate the target sound from the background sound stably, two types of "probe" trials were introduced into the task with relatively low probability (10% of total trials for each type, see chapter 2.3.2 and figure 2.4). "Probe" trials had no reward or light-off given out even when subjects responded within the corresponding response time window. If a probe sound is perceptually close to the background sound, then the subject would not be able to discriminate the probe sound from the background sound in this type of trials, and thus gives a hit rate on this type of "probe" trials that is comparable to the false alarm rate of the same testing session. On the other hand, if a probe sound is perceptually close to the target sound, then the subject would have difficulty telling this type of "probe" trials apart from real "target" trials even after getting extensive experience on the task, and thus gives a hit rate on this type of "probe" trials that is comparable to the hit rate of real "target" trials of the same testing session.

Figure 3.1 shows spectra of these sounds used in the current testing task. "Probe 1" trials had the sound bearing the same periodicity as the "background" sound but without the F0 component presented. "Probe 2" trials had the sound with an F0 component presented just as the "background" sound but bore an increased periodicity that is one semitone higher than the "background" sound.

Figure 3.2 shows hit rate and response latency results from two exemplar subjects, both under the noise masker condition at ~42 dB SPL / ERB. The subject M4Y finished 1180 trials and the subject M61Z finished 770 trials. The numbers of the behavioral performance are shown in table 3.2 and the statistics summary is shown in table 3.3 as well.

Subjects firstly showed a robust discrimination of the target sound from the background sound. False alarm rates were consistently low (<25%) on “sham" trials and hit rates were consistently high (>70%) on “target” trials. The d’ between the target sound and the background sound is higher than 2 (2.15 for subject M4Y, 2.53 for subject M61Z), suggesting a robust discrimination. The average response latency of “target” trials is roughly around one second, whereas the average response latency of “sham” trials is roughly 2.5 seconds. In terms of response rate and latency, “sham” trials are significantly different from “target” trials (response rate: p=2.3e-7 for M4Y, p=3.0e-5 for M61Z; response latency: p=9.6e-14 for M4Y, p=2.7e-5 for M61Z).

“Probe 1” trials bore a probe sound that had the same periodicity with the background sound, but with the absence of an F0 component, just as the target sound. Subjects’ performance on “probe 1” trials was not significantly different from that on “sham” trials (response rate: p=0.067 for M4Y, p=0.42 for M61Z; response latency: p=0.87 for M4Y, p=0.89 for M61Z), but is significantly different from that on “target” trials (response rate: p=9.3e-5 for M4Y, p=0.0017 for M61Z; response latency: p=8.9e-12 for M4Y, p=0.029 for M61Z), suggesting subjects didn’t pay attention to the absence of an F0 component in the probe 1 sound, which is different from the background sound. And the probe 1 sound is thus perceptually not discriminable from the background sound (d’=0.304 for the subject M4Y, d’=-0.326 for the subject M61Z).

“Probe 2” trials bore a probe sound that had the same periodicity with the target sound, but with the presence of an F0 component, just as the background sound. Subjects’ performance on “probe 2” trials is significantly different from that on “sham” trials (response rate: p=1.1e-4 for M4Y, p=0.0040 for M61Z; response latency: p=9.7e-12

for M4Y, p=2.0e-4 for M61Z). And probe 2 sound is clearly discriminable from the background sound (d'=1.943 for the subject M4Y, d'=2.430 for the subject M61Z), suggesting subjects did pay attention to the periodicity change of the probe 2 sound from that of the background sound. Moreover, subjects' performance on "probe 2" trials is not significantly different from that on "target" trials (response rate: p=0.20 for M4Y, p=0.55 for M61Z; response latency: p=0.74 for M4Y, p=0.13 for M61Z), suggesting the probe 2 sound is perceptually close to the target sound.

To sum up, these results suggest (1) subjects can discriminate the target sound robustly from the background sound. (2) "probe 1" sound is perceptually closer to the "background" sound and "probe 2" sound is closer to the "target" sound. (3) Based on (2), the cue that these animals use to discriminate the "target" sound from the "background" sound is based on the periodicity change but not based on whether the F0 component is presented or not. Thus, marmosets don't necessarily need the F0 component to perceive the pitch. They have missing fundamental pitch perception. And this missing fundamental perception is robust at even only one semitone pitch difference, which is the smallest step in Western music melodies.

3.4.2 Effect of noise masker level on missing fundamental perception in marmosets

One possibility to explain the insignificance of the F0 component is that, the noise masker used in the current testing was too high, which may not only mask potential nonlinear distortion product, but may also mask the original F0 component in the background sound. It is interesting to check out whether, and to what extent, the noise masker level influences missing fundamental testing.

Besides playing at a level of ~42 dB SPL / ERB at 440 Hz, the noise masker level was lowered down to levels of ~34 dB SPL / ERB, and ~22 dB SPL / ERB. Results are shown in figure 3.3 and table 3.4 under these three conditions.

False alarm rates were low across all conditions (5.2%, 13.2%, 16.9%, respectively). Whereas, the hit rates of "target" trials were all higher than 75% (82.0%, 88.9%, 79.0%, respectively). The d's between "background" and "target" were all above 1 (2.5429, 2.3382, 1.7666, respectively), suggesting a robust discrimination no matter what noise masker level it was. There was a slight monotonic decrease of d' when the noise level decreases.

The hit rates of "probe 1" trials were all lower than 30% (2.6%, 29.7%, 27.0%, respectively). The d's between "background" and "probe 1" were all lower than 1 (-0.3173, 0.5821, 0.3464, respectively), suggesting the probe 1 sound was not well discriminable from the background sound, and thus the presence or absence of an F0 component was not a used cue no matter what noise masker level it was. There was a non-monotonic change on d'. The hit rates of "probe 2" trials were all higher than 65% (78.2%, 80.6%, 67.5%, respectively). The d's between "background" and "probe 2" were all above 1 (2.4059, 1.9807, 1.4116, respectively), suggesting the periodicity change was a valid cue during the discrimination. There was a slight monotonic decrease of d' when the noise level decreases.

To sum up, even at the lowest noise masker level we tested (~ 22 dB SPL / ERB at 440 Hz), where the F0 component should be ~30 dB above the noise floor, there was still enough evidence to show a robust missing fundamental perception.

3.5. Discussion

In the current study, common marmosets showed perceptual sensitivity to sound periodicity no matter whether there's an F0 component presented or not, in another word, missing fundamental pitch perception, at a pitch difference of only one semitone, the smallest pitch step of Western music melodies. Previous animal studies have only tested missing fundamental perception at a pitch difference no lower than 3 semitones. The precision of the perception we showed here lays a cornerstone for further more music related hypotheses and experimental designs, such as pitch based music melody discrimination, octave generalization, consonance and dissonance perception, etc.

One may argue that the discrimination precision between the probe 2 sound and the background sound may still come from the frequency discrimination of the fundamental frequency component. Pure tone frequency discrimination limens (FDL) were measured in marmosets [Osmanski and Song et al 2016]. Psychometric curves are shown in figure 3.4. FDL of 440 Hz is above two semitones, suggesting that subjects cannot solely rely on the frequency change on F0 component to perform the discrimination. Moreover, the relative FDL in marmosets has a "floor" effect starting around 3.5 kHz (figure 3.5, table 3.5). It is most likely that subjects utilized cues mainly from the kHz spectral range to perform the discrimination. Among primates, the marmoset is not particularly prominent in pure tone frequency discrimination (figure 3.6), Interestingly, non-human primates are not particularly better in pure tone frequency discrimination than other common experimental non-human mammals used in auditory studies (figure 3.7).

Another thing one may argue is the potential possibility that our subjects performed the discrimination by focusing on a local frequency component, instead of using a globally assembled pitch percept to do so. However, in the next chapter we will provide further evidence (chapter 4.5.2) to show this possibility is unlikely to be the case, and our subjects most likely used a harmonicity based cue, and thus pitch, to perform the discrimination.

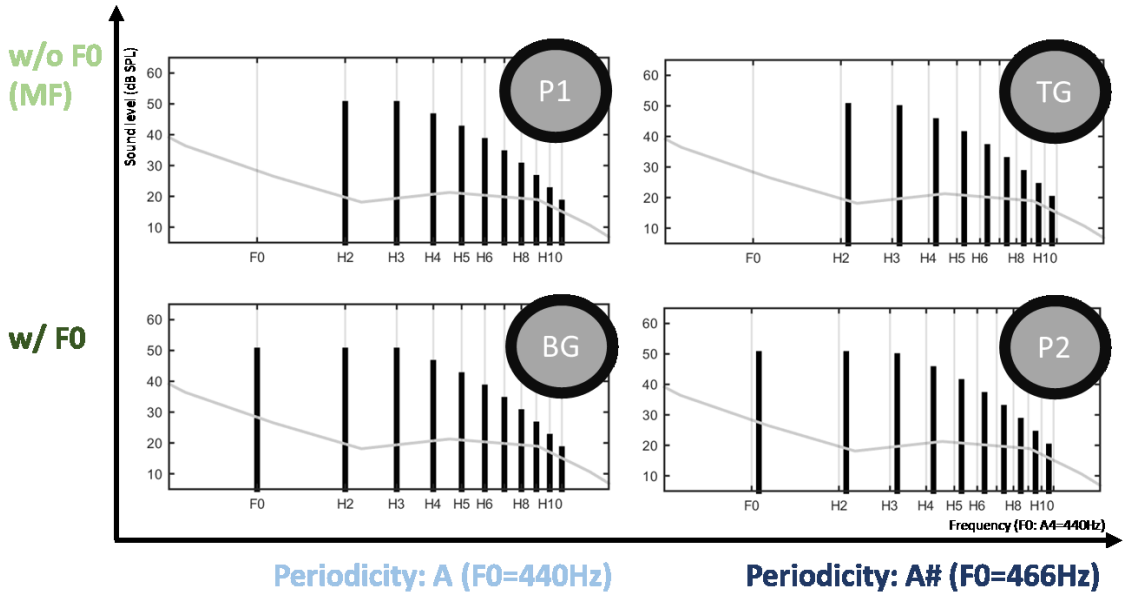


Figure 3.1 Sound stimulus design for testing missing fundamental in marmosets

The background sound (BG), the target sound (TG), the probe 1 sound (P1), and the probe 2 sound (P2) used in the testing procedure of the generalization task. TG was different from BG in two dimensions, sound periodicity and F0 presence. Whereas P1 and P2 were different from BG in only one dimension. P1 and P2 were also different from TG in only one dimension, respectively.

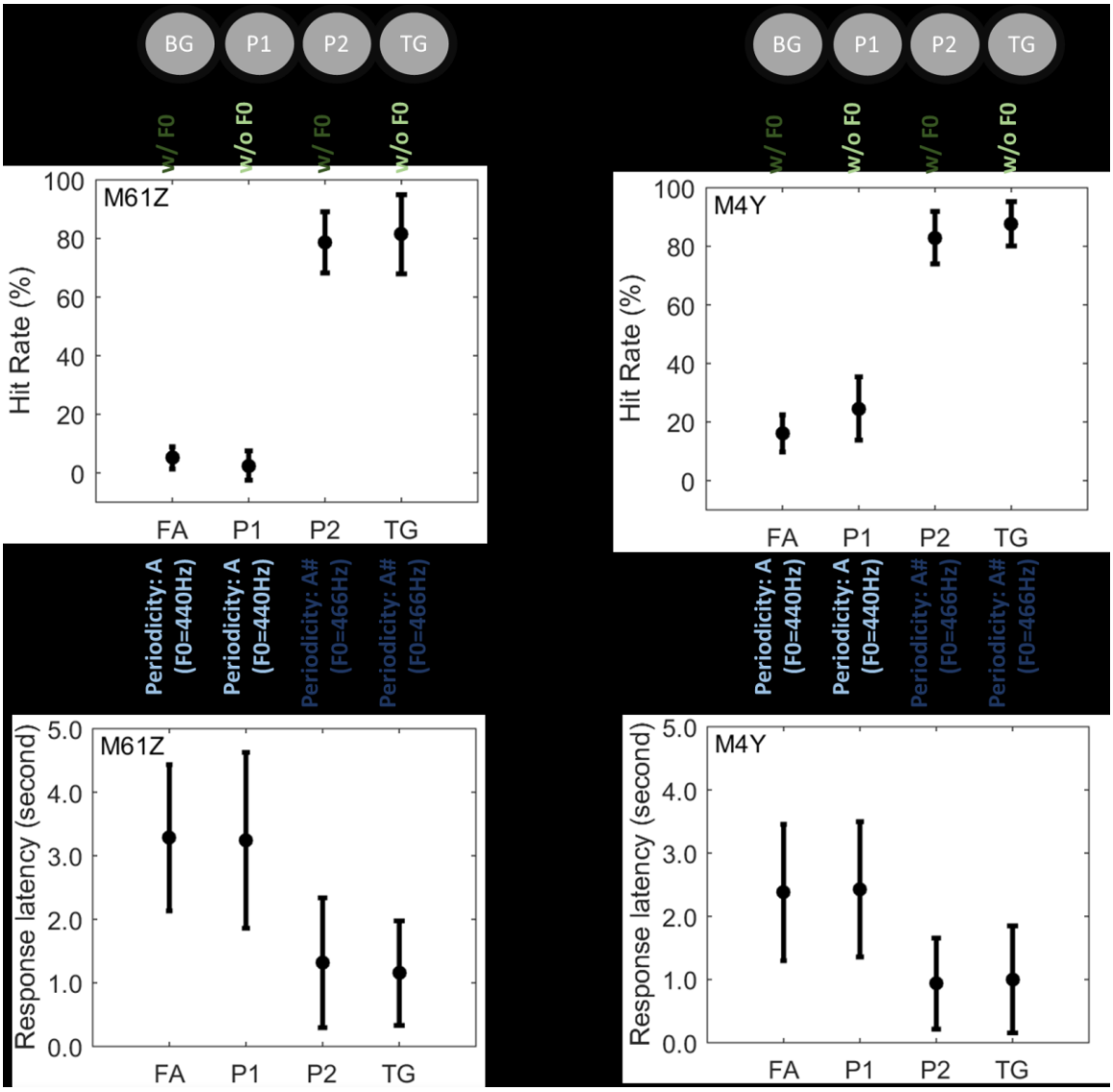


Figure 3.2 Marmoset missing fundamental testing results

Marmoset missing fundamental testing results from the generalization task, with subject M61Z on the left, and subject M4Y on the right. Hit rates are shown in the upper panels. And response latencies are shown in the lower panels.

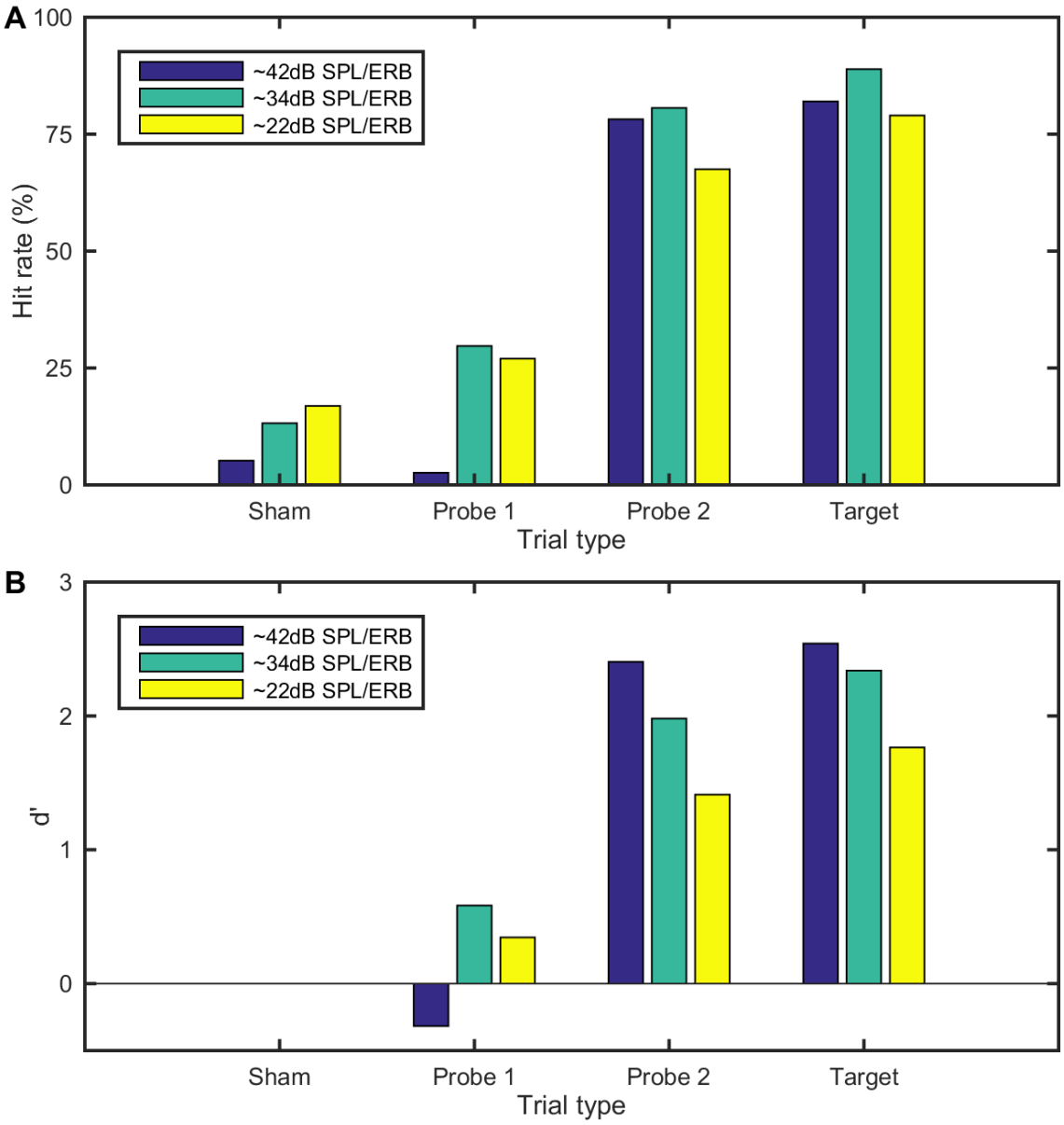


Figure 3.3 Effect of noise masker level on missing fundamental testing

The generalization task was tested under three different noise masker levels (~42 dB SPL/ERB, ~34 dB SPL/ERB, ~22dB SPL/ERB @ 440 Hz). Results shown in hit rates and d'.

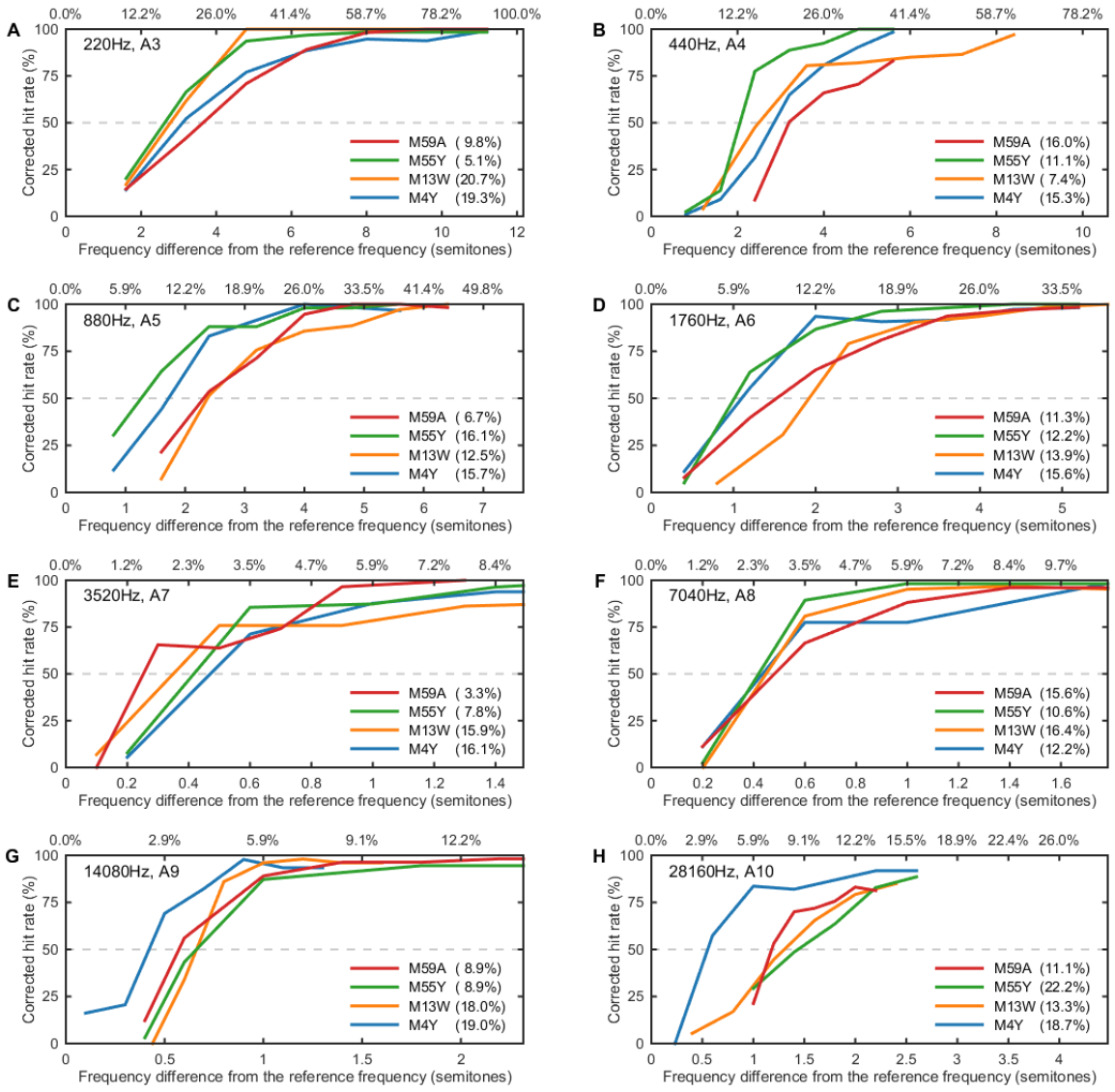


Figure 3.4 Marmoset pure tone frequency discrimination psychometric curves

Representative psychometric curves from marmosets at each reference frequency tested (220 Hz [A3] – 28160 Hz [A10]) [Osmanski and Song et al 2016]. Changes in corrected hit rate are shown as a function of the difference between reference and target frequencies (in both units of semitones and as a percentage change from the reference

frequency). Dashed lines show 50% correct threshold. False alarm rates (measured as a percentage) on each frequency are displayed in the legend next to each subject.

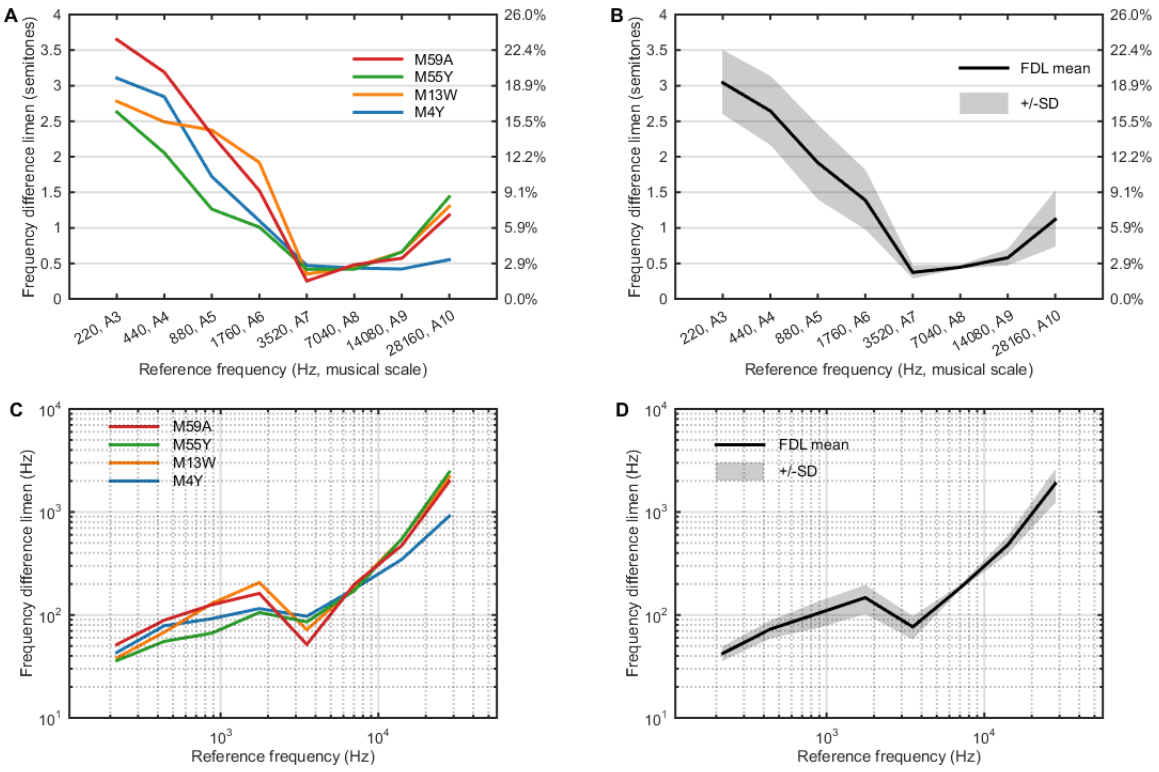


Figure 3.5 Summary of marmoset pure tone FDLs

Relative FDLs (in both units of semitones and as a percentage change from the reference frequency) are shown in panels A (individual data) and in panel B (averaged data, shaded area denotes one standard deviation) [Osmanski and Song et al 2016]. Absolute FDL values are shown in panels C (individual data) and D (averaged data, shaded area denotes one standard deviation).

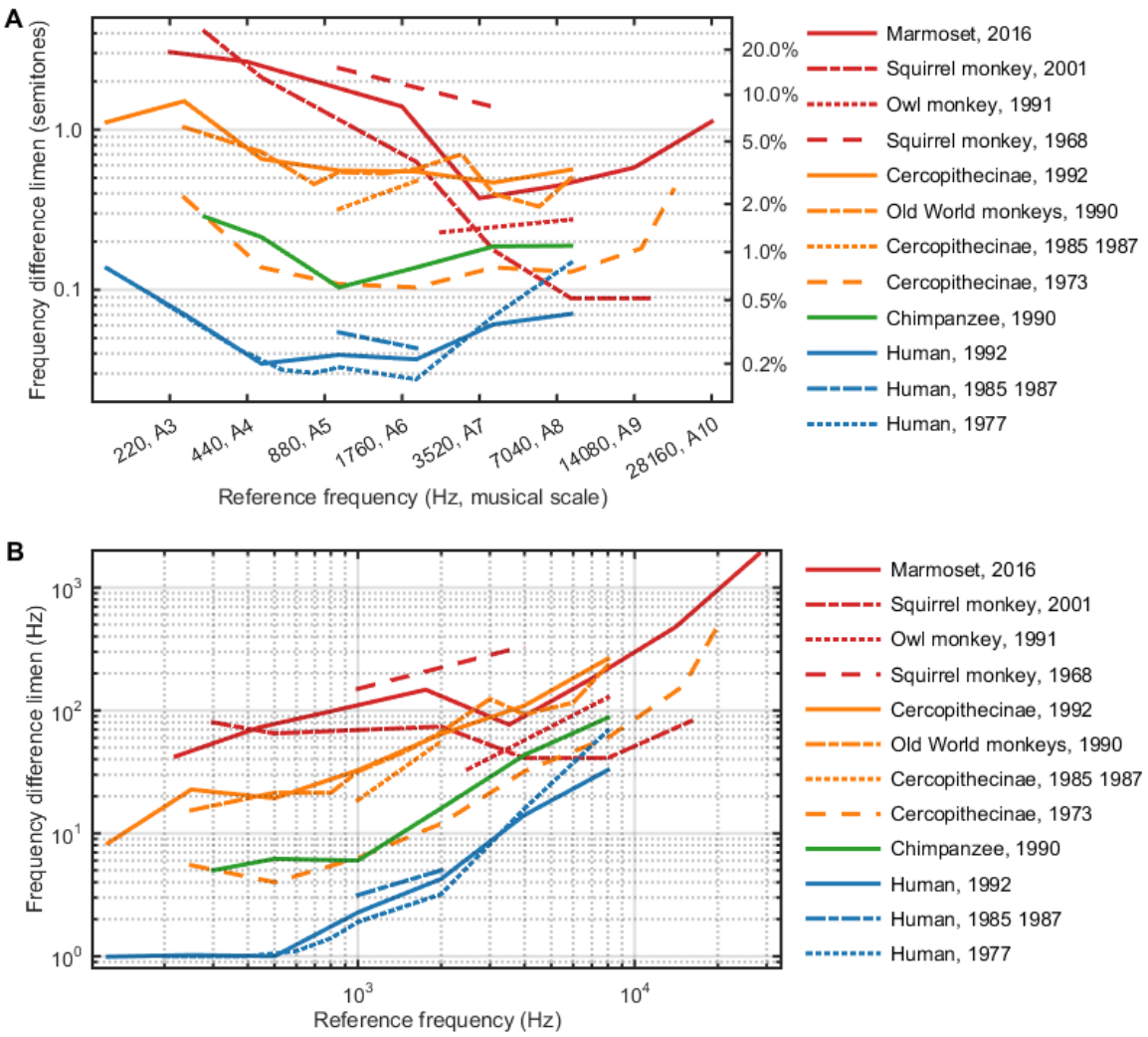


Figure 3.6 Primate pure tone FDL comparison

Comparison of FDLs obtained from marmosets [Osmanski and Song et al 2016] (solid red line) with FDLs of other primate species measured by previous studies (squirrel monkey [Wienicke et al 2001; Capps and Ades 1968], owl monkey [Recanzone et al., 1991], Old World monkeys [Prosen et al 1990], Cercopithecinae [Sinnott et al 1992; Sinnott et al 1987; Sinnott et al 1985; Stebbins 1973], Chimpanzee [Kojima 1990], and human [Sinnott et al 1992; Sinnott et al 1987; Sinnott et al 1985; Wier et al 1977]). In

general, humans show the lowest FDLs, followed by non-human apes and Old World monkeys. New World monkeys show the largest FDLs compared to other primates.

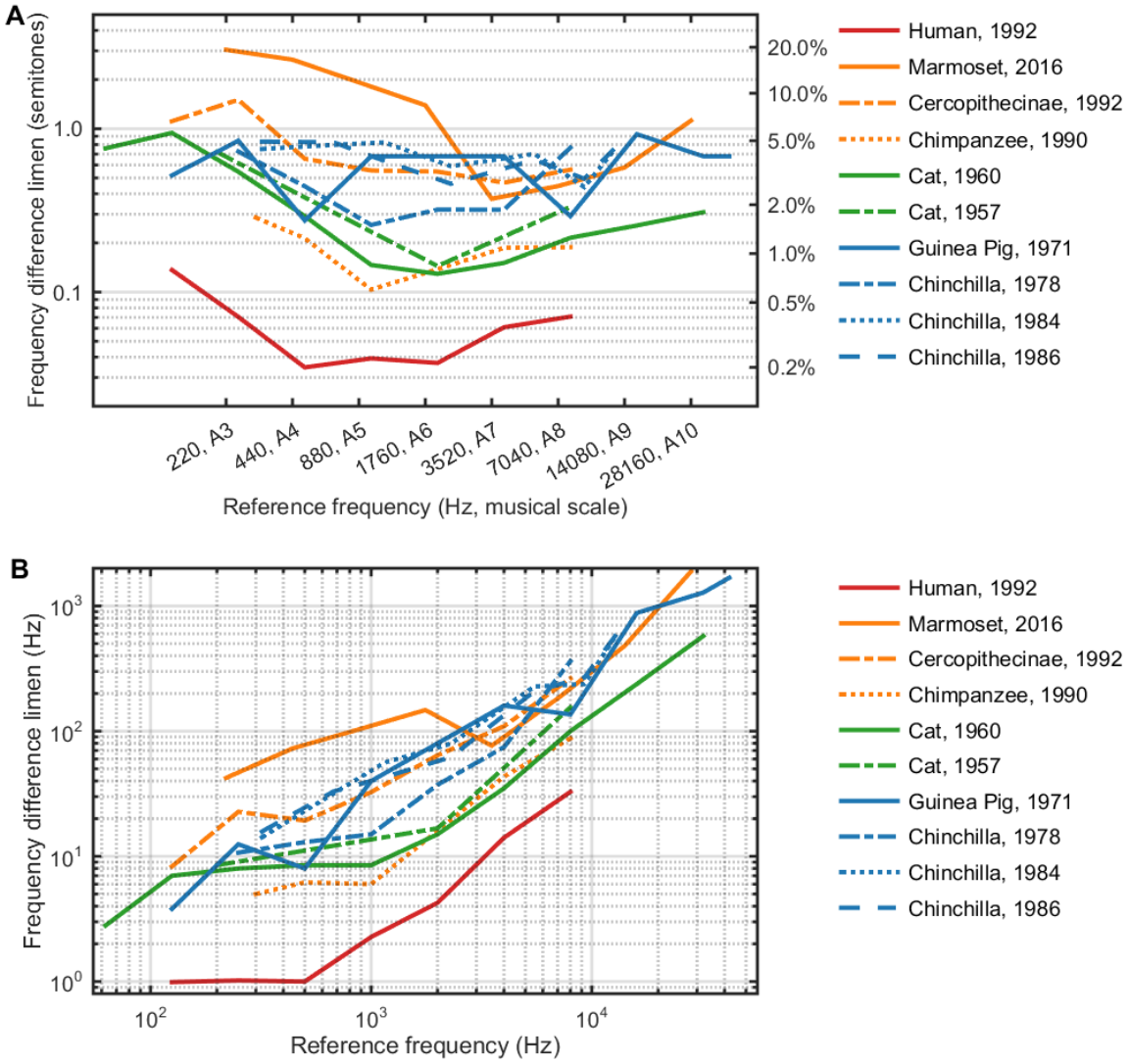


Figure 3.7 Mammal pure tone FDL comparison

Comparison of FDLs obtained from marmosets [Osmanski and Song et al 2016] (solid red line) with FDLs of other mammal species measured by previous studies (Cercopithecinae [Sinnott et al 1992], chimpanzee [Kojima 1990], and human [Sinnott et al 1992], cat [Elliott et al 1960, Butler et al 1957], guinea pig [Heffner et al 1971], chinchilla [Nelson and Kiester 1978, Long and Clark 1984, Clark and Bohne 1986]). In

general, non-human primates are not particularly better in FDL performance than other non-human mammals.

Species	Publication Year, lab	Pitch difference (semitones)	Potential Artifacts: periodicity change (P), distortion product cue (DP), overall sound Level cue (OSL), individual harmonic amplitude (IHA), overall spectrum shape cue (SS), upper most component cue(UMC), lower most component cue (LMC)
Cat	1976, Whitfield 1980, Whitfield	~3 (inharmonic)	P, DP, OSL, SS, UMC, LMC, IHA
Cat	1976, Colavita	7	P, DP, OSL, SS, UMC, LMC, IHA
Songbird	1986, Cynx	8.5	P, DP, OSL, SS, UMC, LMC, IHA
Songbird	2000, Okanoya	8.5	P, DP, OSL, SS, UMC, LMC, IHA
Macaque	1988, Tomlinson & Schwarz	7, 12	P, DP, OSL, SS, UMC, LMC, IHA
Goldfish	2005, Fay	3:3:12, 4:4:12	P, DP, OSL, SS, UMC, LMC, IHA
Chinchilla	2011, Shofner	24	P, DP, OSL, SS, UMC, LMC, IHA

Table 3.1 Summary of animal missing fundamental literatures

A summary of literatures on missing fundamental testing in non-human animals. The smallest pitch change used in the previous studies was 3 semitones.

		"Sham" trials	"Probe 1" trials	"Probe 2" trials	"Target" trials
M4Y	Hit rate or false alarm rate (%)	16.11±6.33 (n=12)	24.63±10.79 (n=6)	82.96±8.97 (n=6)	87.69±7.56 (n=36)
	Response latency (s)	2.38±1.08 (n=38)	2.43±1.07 (n=29)	0.93±0.72 (n=98)	1.00±0.85 (n=621)
	d' from background	-----	0.3037	1.9426	2.1493
M61Z	Hit rate or false alarm rate (%)	5.11±3.79 (n=8)	2.50±5.00 (n=4)	78.68±10.44 (n=4)	81.42±13.50 (n=24)
	Response latency	3.28±1.15 (n=8)	3.24±1.38 (n=2)	1.32±1.02 (n=61)	1.15±0.82 (n=379)
	d' from background	-----	-0.3257	2.4295	2.5278

Table 3.2 Results of missing fundamental pitch testing

The results of missing fundamental pitch testing from two exemplar subjects, shown in hit rate, response latency and d'.

p-value		"Sham" vs "probe 1"	"Sham" vs "probe 2"	"Sham" vs "target"	"Probe 1" vs "probe 2"	"Probe 1" vs "target"	"Probe 2" vs "target"
M4Y	Hit rate	0.0647	1.0774e-04	2.3059e-07	0.0022	9.2645e-05	0.2009
	Response latency	0.8743	9.6539e-12	9.6439e-14	1.2303e-10	8.9497e-12	0.7422
M61Z	Hit rate	0.4242	0.0040	3.0066e-05	0.0286	0.0017	0.5510
	Response latency	0.8889	1.9889e-04	2.7277e-05	0.0573	0.0293	0.1292

p-value > 0.05: not significant
0.05 > p-value > 0.01: *
0.01 > p-value > 0.001: **
0.001 > p-value > 0.0001: ***
0.0001 > p-value: ****

Table 3.3 Summary of statistics results of missing fundamental testing

The summary of p-value statistics of the results shown in table 3.2. Font color indicates statistical significance level. Purple means not significant, whereas brown means significant. The darker the brown is, the more significant the p-value is.

Noise masker level		"Sham"	"Probe 1"	"Probe 2"	"Target"
~42dB SPL/ERB	Hit rate (%)	5.2	2.6	78.2	82.0
	d'	-----	-0.3173	2.4059	2.5429
~34dB SPL/ERB	Hit rate (%)	13.2	29.7	80.6	88.9
	d'	-----	0.5821	1.9807	2.3383
~22dB SPL/ERB	Hit rate (%)	16.9	27.0	67.5	79.0
	d'	-----	0.3464	1.4116	1.7666

Table 3.4 Effect of noise masker level on missing fundamental testing

The effect of noise masker level on the result of missing fundamental testing. No matter what noise masker level it was, the probe 1 sound was always not discriminable from the background sound (d'<1), whereas the probe 2 sound was always discriminable from the background sound (d'>1).

Musical scale		A3	A4	A5	A6	A7	A8	A9	A10
Frequency		220	440	880	1760	3520	7040	14080	28160
FDL (semitones)	M4Y	3.11	2.84	1.72	1.10	0.47	0.43	0.42	0.55
	M13W	2.78	2.49	2.37	1.92	0.35	0.45	0.66	1.30
	M55Y	2.63	2.05	1.26	1.01	0.42	0.42	0.66	1.44
	M59A	3.65	3.19	2.31	1.53	0.25	0.48	0.57	1.18
	mean	3.04	2.65	1.92	1.39	0.37	0.45	0.58	1.12
	SD	0.45	0.49	0.53	0.42	0.09	0.03	0.11	0.39
FDL (%)	M4Y	19.7	17.9	10.5	6.6	2.8	2.5	2.5	3.2
	M13W	17.4	15.5	14.7	11.7	2.0	2.6	3.9	7.8
	M55Y	16.4	12.6	7.6	6.0	2.4	2.5	3.9	8.7
	M59A	23.5	20.2	14.3	9.2	1.5	2.8	3.4	7.1
	mean	19.2	16.5	11.8	8.4	2.2	2.6	3.4	6.7
	SD	3.1	3.3	3.4	2.6	0.6	0.2	0.7	2.4
FDL (Hz)	M4Y	43.2	78.6	92.0	115.5	97.0	178.9	346.8	913.4
	M13W	38.3	68.1	129.3	206.4	71.9	184.7	548.4	2202.5
	M55Y	36.1	55.5	66.6	105.8	85.8	172.5	548.1	2439.1
	M59A	51.6	89.0	125.7	162.1	51.7	198.3	473.2	1987.0
	mean	42.3	72.8	103.4	147.4	76.6	183.6	479.1	1885.5
	SD	6.9	14.4	29.7	46.4	19.5	11.0	95.1	673.8
Testing order	M4Y	3	1	2	4	5	6	7	8
	M13W	4	1	2	3	6	5	7	8
	M55Y	3	2	1	7	6	4	5	8
	M59A	6	5	2	3	1	4	7	8
SL (dB)		40	40	40	40	40	40	40	50

Table 3.5 Marmoset pure tone FDLs

Threshold values (relative FDL, in semitones and % frequency change, and absolute FDL) for each subject is shown at each reference frequency, along with mean and standard deviation (SD). Testing order and sensation level for each reference frequency are also shown [Osmanski and Song et al 2016].

4. MARMOSET PITCH PERCEPTION MECHANISMS

4.1. Summary

The perception of the pitch of harmonic complex sounds is a crucial function of human audition, especially in music and speech processing. Whether the underlying mechanisms of pitch perception are unique to humans, however, is unknown. Based on estimates of frequency resolution at the level of the auditory periphery, psychoacoustic studies in humans have revealed several primary features of central pitch mechanisms (1) pitch strength of a harmonic complex sound is dominated by resolved harmonics; (2) pitch of resolved harmonics is sensitive to the quality of spectral harmonicity; (3) pitch of unresolved harmonics is sensitive to the salience of temporal envelope cues. Here we show, for a standard musical tuning fundamental frequency of 440Hz [ISO 16], the common marmoset (*Callithrix jacchus*), a New World monkey with a hearing range similar to that of humans, exhibits all these three primary features of pitch mechanisms demonstrated in humans. Thus marmosets and humans may share similar pitch perception mechanisms, suggesting that these mechanisms may have emerged early in primate evolution.

4.2. Introduction

Extracting pitch from periodic complex sounds is one of the most fundamental functions of the human auditory system, and this periodicity is a critical aspect of music, speech, animal vocalizations, stream segregation and many other auditory functions.

Indeed, almost all naturally occurring, pitch-bearing sounds are periodic, and perceptual sensitivity to acoustic periodicity has been demonstrated in a wide variety of vertebrate species – including anurans [Fay 2005], songbirds [Cynx and Shapiro 1986, Okanoya 2000], and mammals [Heffner and Whitfield 1976, Chung and Colavita 1976, Tomlinson and Schwarz 1988, Shofner 2011]. The perceptual mechanisms used by humans to extract a sound's pitch have been extensively studied, but to date, there has been little evidence to suggest that any other species use mechanisms similar to those found in humans [Plack et al 2005, Shofner and Chaney 2013]. Further, frequency resolution at the level of the auditory periphery, an important component of pitch perception, is thought to be insufficient in other mammalian species to produce human-like pitch perception [Shofner and Chaney 2013, Klinge et al 2010, Shera et al 2002, Shera et al 2010]. This lack of adequate frequency resolution limits comparisons of central mechanisms with other species using the same analytical criteria established for humans.

Recent data from two macaque monkey species, representing Old World primates, and the marmoset, a highly vocal New World primate species separated from humans by about 30 to 40 million years [Worley et al 2014] and phylogenetically located roughly between macaques and other non-primate mammals tested in pitch studies, have begun to cast doubt on whether these pitch perception mechanisms are unique to humans. Both physiology data from the macaque monkey [Joris et al 2011] and behavioral data from marmosets [Osmanski et al 2013] suggest that these primate species may exhibit frequency resolution at the auditory periphery more similar to that seen in humans than that seen in previously tested mammals. Based on these findings, I hypothesized that there may also be perceptual mechanisms for pitch perception shared between humans

and other primates. In this chapter, evidence is shown that marmosets exhibit all primary features of central pitch mechanisms that have been demonstrated in humans. Based on these analyses, we suggest that these central pitch perception mechanisms are not unique to humans, but can likely be found in non-human primates - including New Word primate species - and thus may have originated relatively early in primate evolution.

Most pitch-evoking sounds occurring in natural environments have spectra with harmonic structure, in which the acoustic power is concentrated at frequencies that are integer multiples (harmonics) of a common fundamental frequency (F0). These harmonics are processed at the cochlea by a bank of peripheral auditory filters which separate the incoming signal into individual frequency channels along a tonotopic axis [Plack et al 2005]. The tuning bandwidth of these filters increases, and thus the frequency resolving power decreases, as frequency increases [Glasberg and Moore 1990, Moore 2012]. In humans, only the lowest 5~10 harmonics are well segregated into different auditory filters, and can be heard individually from the whole complex sound [Plomp 1964, Plomp and Mimpen 1968]. These harmonics are known as resolved harmonics (RES). The tuning bandwidth of auditory filters at high frequencies becomes larger than the spacing of adjacent harmonics, which is equal to F0 in a pitch-evoking sound. Thus each auditory filter receives significant power from more than one harmonic, and these are defined as unresolved harmonics (URS). The tuning bandwidths in marmosets were previously measured [Osmanski et al 2013]. For an F0 of A440 according to the musical tuning standard (440Hz), based on these data, a model of the excitation pattern at the level of the auditory periphery [Moore and Glasberg 1983] was applied to marmosets

(shown in figure 2.10). The model shows distinct peaks at each harmonic for RES and a raised smooth plateau across frequency for URS.

In humans, the upper boundary of RES and the lower boundary of URS can be assessed behaviorally [Plomp 1964, Plomp and Mimpen 1968, Shackleton and Carlyon 1994, Bernstein and Oxenham 2006] and these measures can be used to determine the relationship between bandwidth and F0 at these two boundaries [Moore 2012, Shackleton and Carlyon 1994, Bernstein and Oxenham 2006], as described in chapter 2.5 (figure 2.11 (A), dashed black lines). We applied these boundary ratios derived in humans to tuning bandwidths measured in marmosets [Osmanski et al 2013] to estimate resolvability boundaries for marmosets (figure 2.11 (A), solid black lines). Noticeably, for an F0 of 440Hz, marmosets appear to have as many RES as humans.

4.3. Primary Features of Human Complex Sound Pitch Perception

4.3.1 Dominance in pitch strength

Given the functional properties of the auditory periphery outlined above, there are several possible central mechanisms which can theoretically be used to extract cues to decode the pitch of a harmonic sound (see chapter 1.3 as well). The simplest mechanism is to use the lowest frequency component present in a harmonic sound to extract pitch, which is usually equal to a pure tone at F0 (PTF0) [Ohm 1843, Helmholtz 1863]. Alternatively, the central auditory system may contain many spectral harmonic templates. A match between one of these templates and the RES components of an incoming harmonic complex sound determines the pitch [Goldstein 1973, Shamma and Klein 2000]. A third potential mechanism is to extract pitch from the interactions among URS

components within an auditory filter, which generate a temporal envelope with a periodicity equal to the pitch [Schouten 1938]. Although each of these mechanisms alone is sufficient to evoke pitch, the relative strengths of the perceived pitch based on these mechanisms are different. Pitch strength (a measure of the salience of the perceived pitch) is commonly believed to be correlated with the ability to discriminate changes in F0 [Micheyl et al 2010]. The smallest change in F0 that a subject can discriminate, measured as the F0 difference limen (F0DL), is thought to be inversely related to pitch strength. Previous F0DL studies have revealed that harmonic complex tones have a greater pitch strength than a pure tone at the same F0 [Zeitlin 1964, Henning and Grosberg 1968, Fastl and Weinberger 1981, Spiegel and Watson 1984], implying that the PTF0 alone cannot dominate the pitch strength of complex tones. Indeed, removing the F0 component from a harmonic complex tone doesn't affect the robustness and the strength of pitch perception (the phenomenon of the "missing fundamental") [Schouten 1938, Thurlow and Small 1955, Licklider 1956, as well as chapter 3]. Further, a significant F0DL increase (and consequently a decrease in pitch strength) was reported as spectral content was shifted out of the lower RES range [Houtsma and Smurzynski 1990, Kaernbach and Bering 2001], suggesting that RES have a greater pitch strength than URS and thus dominate the pitch strength of harmonic complex tones when both RES and URS are present. This is the first primary feature of human pitch perception mechanisms.

4.3.2 Spectral harmonicity pitch on resolved harmonics

The second primary feature of human pitch perception mechanisms is that the pitch of RES is sensitive to the fidelity of spectral harmonicity. The F0DL of RES could

reflect sensitivity to a globally assembled pitch or, alternatively, to local changes to the individual components within each auditory filter [Faulkner 1985]. Importantly, sensitivity to individual components would not necessarily require them to be in a harmonic relationship. A series of experiments that measured the change in F0DLs for inharmonic versus harmonic sounds found, where both groups had the same average resolvability and spectral range, F0DLs were higher under the inharmonic condition [Moore and Glasberg 1990, Micheyl et al 2010, Micheyl et al 2012]. These results have been taken as evidence for a role of harmonicity in lowering F0DLs and, since F0DL for RES is affected by changes to the fidelity of harmonicity, suggest that F0DL reflects sensitivity to a globally assembled pitch.

4.3.3 Temporal envelope pitch on unresolved harmonics

The third primary feature of human pitch mechanisms is that the pitch strength of URS is related to the salience of temporal envelope cues. Varying phase relationships among individual components of a harmonic complex sound can induce temporal envelope changes without affecting spectral amplitude. For example, a harmonic complex in Schroeder phase [Schroeder 1970] has a flattened temporal envelope, and thus a reduction in the salience of temporal envelope cues, yet retains the same spectral profile as the same harmonic complex in sine phase. Harmonic complex sounds in Schroeder phase show increased F0DL compared to those in sine phase for URS, but not for RES [Houtsma and Smurzynski 1990]. Thus changes in temporal envelope cues only affect the perceived pitch of URS.

4.3.4 Lack of human-like primary features in nonhuman mammals

These three primary features of human pitch perception are summarized in figure 4.1. Among these three primary features, it has been shown that other mammals like chinchillas [Shofner and Chaney 2013] and macaque monkeys [Joly et al 2014] can perceive sound periodicity through temporal processing. However, to the best of our knowledge, for any non-human species, neither the relative importance of RES in pitch strength, nor the sensitivity to spectral harmonicity in RES pitch has been shown behaviorally. The lack of evidence for these two features in animals has led to the proposal that pitch perception arises solely from temporal, rather than spectral, processing in non-human mammals [Shofner and Chaney 2013].

4.4. Methods

4.4.1 Subjects, tasks, and acoustic stimuli

The behavioral apparatus, threshold measuring task, and related analysis methods have been discussed in chapter 2. In this part of the study. Four subjects were tested. Animals were trained to detect the appearance of targets sounds, which had “F0s” that were always higher than 440Hz, from “background” sounds, which had an “F0” equal to 440Hz, also known as A440 or A4 in musical tuning standard [ISO 16]. A sample psychometric curve from raw hit rates is shown in figure 2.2 on subject M13W under an unresolved harmonics condition.

To test the primary feature #1 of human-like pitch perception mechanisms, animals were tested under four conditions: pure tone at fundamental frequency alone (PTF0), all harmonics presented together (ALL), resolved harmonics (RES), and

unresolved harmonics (URS). Resolved und unresolved harmonics boundaries were estimated in chapter 2.5.3.

For PTF0 discrimination, the background level was calibrated to be around 40dB SL (~70dB SPL). Targets were adjusted in level to match the sensation level of the background, based on the marmoset audiogram [Osmanski and Wang 2011] to eliminate level differences as a potential cue. For the other stimuli, the maximum level of harmonics was calibrated to be around 50dB SPL.

Previous results in chinchillas showed that the F0DL of complex tones composed of the first 10 harmonics was lower than the F0DL of a pure tone of the same F0 [Shofner 2000]. However, these results cannot exclude the possibility that discrimination in that task was based on the relative location of the highest harmonic on the spectral edge alone rather than discriminating pitch per se [Nelson and Kiester 1978]. To minimize the possibility that our subjects could use spectral edges as a cue for discrimination, we implemented roll-offs on the spectral edges. Upper spectral edges of RES sounds were rolled off starting from 50dB with a slow 4dB/ 440Hz slope as *Level = (50 - (f – 6 · 440Hz) · 4 / 440Hz) dB SPL*, ending at 7040Hz. Upper spectral edges of ALL sounds, and both edges of URS sounds were rolled off as *Level = (50 + 20 * log10((10^(1-|f-* F_{edge} *|/660)-1)/9)) dB SPL*, where upper F_{edge} = 28.16kHz and lower F_{edge} = 11.88kHz. These spectral envelope roll-off designs were previously used in humans to minimize spectral edge cues [Moore and Moore 2003].

To test the primary feature #2 of human-like pitch perception mechanisms, two additional F0DLs were measured under inharmonic conditions. Inharmonic conditions were generated by introducing inharmonic spectral shifts to the original harmonic RES

sounds, but keeping average spacing between adjacent spectral components as the same as the original harmonic RES. The inharmonic shift condition #1 was used in [Moore and Glasberg 1990], which shifted the odd numbered harmonics up by 15% of its original F0, and even numbered harmonics down by 15% of its original F0. The inharmonic shift condition #2 was used in [Micheyl et al 2010, Micheyl et al 2012], which shifted every harmonic up by 25% of its original F0. Figure 4.6 (A) shows background sounds used under these three conditions: original harmonic RES, inharmonic shift condition #1, and inharmonic shift condition #2. The "F0" of inharmonic conditions is defined as the average spacing between adjacent spectral components, thus these three background sounds have the same F0, which is 440 Hz. Inharmonic F0DLs were measured from each animal under each inharmonic shift condition, where target sounds are shifted following the same shifting paradigm as the background sound, according to their own F0s. The upper spectral roll-off envelope was always fixed for all RES harmonic or inharmonic conditions, as introduced in the previous paragraph, no matter what F0 it was or what harmonic/inharmonic condition it was under.

To further test the primary feature #3 of human-like pitch perception mechanisms. Schroeder phases with a negative sign were generated for both RES and URS. Schroeder phases were designed to minimize the waveform peak amplitude while remaining the spectrum amplitude and power unchanged [Schroeder 1970]. It is believed that Schroeder phases with a negative sign have a flatter temporal envelope compared to Schroeder phases with a positives sign, after phase dispersion in the inner ear [Kohlrausch and Sander 1995]. Schroeder phase sounds have the same spectral envelope as their sine phased counterparts.

For URS measurements, both sine phased and Schroeder phased conditions, a fixed level band-pass white noise was generated online to prevent subjects from using potential non-linear distortion products on the lower frequency side to do the discrimination. The cut-off frequencies of the noise masker were 100 and 12000 Hz. Noise was estimated as 40~46dB SPL / ERB.

Two of the subjects (M13W and M4Y) were implanted with a head-cap designed for neurophysiological experiments [Lu et al 2001], and were head-fixed during all testing sessions.

4.4.2 Data analysis

Qualified experimental sessions (as defined in chapter 2.3.1) from the same subject under the same condition were combined together in temporal order and then equally divided into two analysis parts or measures (first measure and second measure). F0DLs were calculated for each measure. Each measure contains at least 24 repetitions for each target. For comparisons of F0DLs, ratios were calculated between F0DLs, Wilcoxon signed rank test was used to test statistical significance.

To test whether marmosets also exhibit these primary features of human pitch mechanisms, we measured F0DLs in marmosets under eight different conditions. All subjects finished all 8 testing conditions except subject M62U, who had persistent high false alarm rate (>30%) under Schroeder phase URS condition, and was consequently excluded from Schroeder phase RES testing. M62U's Schroeder phase URS data are thus also excluded from analyses in chapter 4.5, but are included in current chapter 4.4 for

method comparison purpose. Testing order on each subject was listed in table 4.1. The F0DL and false alarm rate values were calculated and listed in table 4.2.

To assure that the measured F0DLs were stable over time, stability ratios of the two measures (2^{nd} / 1^{st}) was calculated under each condition for each animal (figure 4.2). F0DL stability ratios, defined as the ratio between the second F0DL and the first F0DL measured on the same animal and under the same condition. F0DL stability ratios based on corrected hit rate were plotted to the left, with the mean value (1.03), SEM (1.03), SD (1.18). There was no significant difference between the two F0DL measures (p=0.610, Wilcoxon signed-rank test, n=31). F0DL stability ratios based on d' were plotted in the middle, with the mean value (1.02), SEM (1.06), SD (1.36). There was no significant difference between the two F0DL measures (p=0.754, Wilcoxon signed-rank test, n=31). False alarm rate stability ratios were plotted to the right, with the mean value (0.92), SEM (1.08), SD (1.55). There was no significant difference between the two false alarm rate measures (p=0.249, Wilcoxon signed-rank test, n=31). Both F0DLs and false alarm rates were stable in our current dataset.

4.4.3 Robustness of F0DL calculation

F0DL can be defined and calculated either based on corrected hit rate, or based on d' (see chapter 2.3). Both calculations take not only hit rate but also false alarm rate into account. Derived from the signal detection theory, d' based F0DL is mathematically more rigorous comparing to corrected hit rate based F0DL calculations. In practice, it is interesting to see whether and how these calculations differ based on the currently acquired dataset.

The figure 4.3 (A) shows the ratio between the F0DL based on d' and the F0DL based on corrected hit rate of each measure. The mean of the ratio is 0.85, with SD of 1.22, SEM of 1.03. F0DLs based on d' are significantly smaller from F0DLs based on corrected hit rate ($p=6.4e-11$, Wilcoxon signed rank test, one side, $n=62$). This result suggests F0DL calculation based on d' are systematically lower than F0DL calculation based on corrected hit rate in our dataset.

To test whether the disagreement between the two F0DL calculations is at least partially due to the variation of false alarm rate, the ratio between two different F0DL calculations were plotted against the false alarm rate of the same measure in figure 4.3 (B). A linear regression model shows the fitted ratio as $1.526 \cdot (\text{False alarm rate}) + 0.623$ ($R^2=0.611$, adjusted $R^2=0.604$, and $p=6.7e-14$ for the linear coefficient), as indicated by the grey dashed line. The model shows the variation of the ratio is largely dependent on false alarm rate. And when the false alarm rate is around 24.7%, the ratio is around 1, where the two F0DL calculations give the same numerical estimation.

Since the ratio is largely dependent on the false alarm rate. It's possible that this dependence came from the divisor (d' based F0DL), or the dividend (corrected hit rate based F0DL), or both. Two measures were made for each animal under each condition, and both F0DL stability ratio and false alarm rate stability ratio were calculated, as shown in figure 4.1, the F0DL stability ratio can be plotted against false alarm rate stability ratio to see whether and how F0DL changes while false alarm rate increases, for both F0DL calculations. Figure 4.3 (C) and (D) show the results on both corrected hit rate based F0DL and d' based F0DL. Corrected hit rate based F0DL does not change significantly while false alarm rate changes (linear regression model, F0DL ratio = -0.016

· (False alarm rate ratio) + 1.055, R^2=0.0016, adjusted R^2=-0.0329, and p=0.832 for the linear coefficient), whereas d' based F0DL increases significantly while false alarm rate increases (linear regression model, F0DL ratio = 0.5202 · (False alarm rate ratio) + 0.5492, R^2=0.2953, adjusted R^2=0.2710, and p=0.0016 for the linear coefficient). This analysis suggests the d' based F0DL calculation is sensitive to the false alarm rate change, whereas corrected hit rate based F0DL calculation is not sensitive to the false alarm rate change.

Since one of the criterion for a qualified experimental session of our current dataset is that an experimental session's false alarm rate should be lower than 25%, it is possible that our criterion selection is favorable for corrected hit rate based F0DL calculation. Based on the results shown in figure 4.3 (B), if the animal's condition can be maintained within a small range around ~ 25% false alarm rate, it is interesting to see how these two calculations differ from each other again.

Several outliers are noticeable in d' based F0DL figures. The uppermost outlier of the middle column in figure 4.2, the lowermost outlier in figure 4.3 (A), the lower left corner outlier in figure 4.3 (B), and the upper right corner outlier in figure 4.3 (C) are all due to a single measure: the first measure of Schroeder phase URS in subject M11X, which shows a reasonable corrected hit rate based F0DL (1.300 semitones), but a dramatically decreased d' based F0DL (0.325 semitones) (see table 4.2). As the measure's false alarm rate is low (3.5%), which gives a Z-score of -1.81, the raw hit rate on the first target was around 30%, which would be sub-threshold in the corrected hit rate based F0DL calculation, but jumps to be supra-threshold in the d' based F0DL (Z-score ~-0.51, d' = 1.30). When the false alarm rate or the hit rate on the first target is low, Z-

score is more sensitive to a single "hit" event, due to limited statistical power. This measure is further discarded from the d' based F0DL analysis in the following analysis.

4.5. Results

4.5.1 Pitch strength of a harmonic tone is dominated by resolved harmonics

In order to test whether RES dominate pitch strength in marmosets, we measured F0DLs under four stimulus conditions: 1) All harmonics covering the marmoset hearing range of a common F0 (ALL); 2) PTF0; 3) RES only; 4) URS only (figures 4.4 / 4.5 A and B). Example psychometric curves from one marmoset (M13W) are shown in figure 4.4 (C) and figure 4.5 (C) (figure 4.4 is corrected hit rate based F0DL analysis, whereas figure 4.5 is d' based F0DL analysis). This animal can easily discriminate frequency changes greater than one semitone under both ALL and RES conditions, showing hit rates above 75% or d' above 1.5. F0DLs from each condition and across all tested animals are shown in figure 4.4 (D) and figure 4.5 (D). Average F0DLs in four stimulus conditions are 0.51 / 0.41 (ALL), 2.68 / 2.40 (PTF0), 0.37 / 0.31 (RES) and 0.94 / 0.80 (URS) semitones for corrected hit rate based F0DL / d' based F0DL (the same below), respectively. The average pure tone threshold measured in marmosets (440 Hz, 16.7% / 14.9% F0 change) is comparable to reported pure tone thresholds at a similar frequency in squirrel monkeys, another New World monkey species (500 Hz, 13.0%) [Wienicke et al 2001]. F0DL dominance ratios are (defined and shown in figure 4.4 (E) and figure 4.5 (E)) significantly lower than 1 for both PTF0 ($p=0.0039$ / $p=0.0039$) and URS ($p=0.0039$ / $p=0.0078$). The dominance ratios for RES, however, are significantly higher than 1

(p=0.0039 / p=0.0039), suggesting that RES play a critical role in producing F0DL of ALL and likely dominate pitch strength of harmonic complex sounds in marmosets.

4.5.2 Pitch of resolved harmonics is sensitive to the quality of spectral harmonicity

To test whether F0DL of RES in marmosets is sensitive to the fidelity of spectral harmonicity, we followed the same rationale as in human studies and sought to demonstrate that F0DLs of RES in marmosets can be increased using inharmonic shifts. We measured additional F0DLs of modified RES in which spectral components were inharmonically shifted using two methods based on previous human studies. First, we shifted odd numbered harmonics upward by 15% of F0, and even numbered harmonics downward by 15% of F0 (Shift condition #1) [Moore and Glasberg 1990]. Second, we shifted all RES components upward by 25% of F0 (Shift condition #2) [Micheyl et al 2010, Micheyl et al 2012]. Figure 4.6 (A) illustrates these spectral shifts. Generally, any inharmonic shift should produce a more ambiguous pitch compared to the harmonic condition [Micheyl et al 2010]. Figure 4.6 (B, C) shows that Shift condition #1 generated significantly larger F0DLs than the harmonic F0DLs under corrected hit rate based calculation (p=0.0039), which is consistent with findings in humans [Moore and Glasberg 1990], and a near significant trend for increased F0DLs under d' based calculation (p=0.0547). Shift condition #2 showed a trend for increased F0DLs, although not significantly higher than the harmonic F0DLs by itself alone (p=0.191 / p=0.0742, but see explanations in figure 4.7). Together, these results show that inharmonic F0DLs are significantly higher than harmonic F0DLs in marmosets (figure 4.6 (B) and (C), shift conditions #1 and #2 combined, p=0.0081 / p=0.0140), suggesting that spectral

harmonicity is required to achieve a lower F0DL. Thus, F0DL of RES reflects discrimination of a globally assembled pitch in marmosets.

4.5.3 Pitch of unresolved harmonics is sensitive to the salience of temporal envelope cues

Finally, we sought to determine the role of temporal envelope cues in URS-induced pitch in marmosets. We used the Schroeder phase to introduce a flattened temporal envelope on both RES and URS (figure 4.9 (A) and (B), but also see figure 4.8). The Schroeder phase URS condition was much more difficult for one of our subjects, who failed to produce a comparable F0DL due to a persistent high false alarm rate. For the remaining three animals, the Schroeder phase did not introduce a significant F0DL increase for RES ($p=0.422$ / $p=0.422$), but did so for URS ($p=0.016$ / $p=0.031$) (figure 4.9 (C) and (D)), suggesting that marmosets, like humans, are sensitive to the salience of temporal envelope cues of URS. That is, marmosets appear to discriminate periodicity changes on URS in a manner similar to that of human subjects [Houtsma and Smurzynski 1990].

4.6. Discussion

The findings described above show that marmosets share three primary features of pitch perception mechanisms demonstrated in humans (figure 4.1): (1) both species have a higher pitch strength for RES compared to either URS or PTF0 (figure 4.5). Thus pitch strength of a harmonic complex sound is dominated by RES; (2) both species are sensitive to changes in the fidelity of spectral harmonicity for RES components (figure

4.6). (3) both species are sensitive to the salience of temporal envelope cues for URS components (figure 4.9). The first two of these features have been previously demonstrated only in humans and were not believed to be a component of auditory perception of non-human mammals [Shofner and Chaney 2013]. The present data are thus the first to show that a non-human species shares all three primary features of central pitch processing mechanisms with humans. The majority of previous work with non-human species has been done in rodents (e.g., chinchilla [Shofner and Chaney 2013], gerbils [Klinge et al 2010]) and has generally shown an impoverished peripheral frequency resolution, which called into question the existence of human-like pitch perception mechanisms in these species. It was suggested, for example, that chinchillas, unlike humans, may rely solely on temporal envelope cues for perceiving periodicity [Shofner and Chaney 2013].

Rodents share a common ancestor with primates approximately 90 million years ago whereas the separation of New World and Old World monkeys occurred only about 40 million years ago [Worley et al 2014]. Importantly, evidence from behavioral studies in a New World primate, the common marmoset [Osmanski et al 2013] and physiological studies in Old World monkeys [Joris et al 2011] suggests that both of these primate groups at least share similar peripheral frequency resolution with humans.

In addition to these perceptual data, a putative cortical pitch center has been described in marmoset auditory cortex [Bendor and Wang 2005], at the anterolateral low-frequency border of primary auditory cortex, which contains neurons responsive to pitch-evoking sounds in humans. Depending on harmonic resolvability, these neurons extract pitch using either spectral harmonicity or temporal envelope cues [Bendor et al 2012],

and thus mirror features (2) and (3) of the pitch perception mechanisms mentioned above. In humans, correspondingly, sensitivity to pitch strength and harmonic composition has also been reported near the same functional cortical location as identified in marmoset auditory cortex [Norman-Haignere et al 2013, Penagos et al 2004], suggesting a homologous cortical pitch processing center shared by humans and marmosets. In sum, all of these data suggest that the marmoset is a valuable model system to study the neuronal circuitry underlying human-like pitch perception mechanisms and related auditory attributes.

Marmosets have a rich vocal repertoire that contains a variety of harmonic structures. Some of their vocalizations (e.g., "phee", and "twitter" calls) contain high-frequency F0s (>2–3 kHz), whereas others (e.g., "egg," "moan," and "squeal" calls) contain low-frequency F0s (< 2 kHz, in the range of pitch) [Epple 1968, Bezerra and Souto 2008, Agamaite et al 2015]. In addition, harmonic structures are also commonly found throughout the marmoset's natural acoustic environment in the South American rainforest, including the vocalizations of various heterospecific species such as insects, birds, amphibians, mammals, and so forth. Marmosets thus can, and likely do, make use of pitch perception mechanisms in the roles of both predator and prey in their natural habitat as they interact with these other species.

On the basis of these findings, we suggest that human-like pitch perception mechanisms may have originated relatively early in primate evolution, perhaps as early as or even earlier than ~40 million years when New World and Old World primates separated on the evolutionary tree. To more precisely localize the evolutionary origin of pitch perception, more species, including both primate and non-primate species, need to

be tested in a manner similar to what has been conducted in humans and marmosets. The resultant dataset would allow us to more fully describe the evolutionary development of peripheral frequency resolution and central processing mechanisms which have resulted in human-like pitch perception.

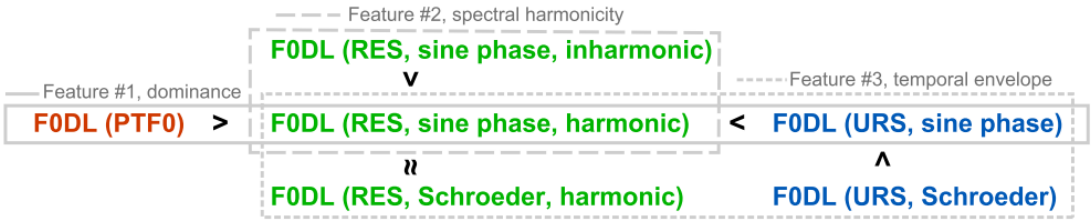


Figure 4.1 Summary of human pitch perception mechanisms

(1) Lower resolved harmonics (RES) have a stronger pitch strength, thus a smaller F0 discrimination threshold, compared to both pure tones at the fundamental frequency (F0) and higher unresolved harmonics (URS); (2) Pitch perception based on resolved harmonics is sensitive to the fidelity of spectral harmonicity; (3) Pitch perception based on unresolved harmonics is sensitive to the saliency of temporal envelope cues.

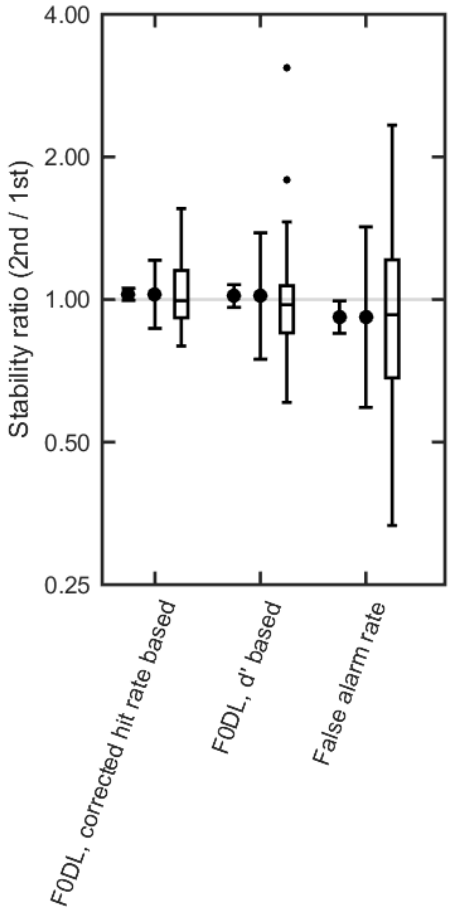


Figure 4.2 Stability of measures

Stability of measurements, indicated in relative ratios between the second measure and the first measure. Each group of ratio statistics shows the mean value with standard error mean (SEM) to the left, the mean value with standard deviation (SD) in the middle, and the box plot to the right.

F0DL stability ratios, defined as the ratio between the second F0DL and the first F0DL measured on the same animal under the same condition. The group of F0DL stability ratios based on corrected hit rate were plotted to the left, with the mean value

(1.03), SEM (1.03), SD (1.18). There was no significant difference between the two F0DL measures (p=0.610, Wilcoxon signed-rank test, n=31). The group of F0DL stability ratios based on d' were plotted in the middle, with the mean value (1.02), SEM (1.06), SD (1.36). There was no significant difference between the two F0DL measures (p=0.754, Wilcoxon signed-rank test, n=31). The group of false alarm rate stability ratios were plotted to the right, with the mean value (0.92), SEM (1.08), SD (1.55). There was no significant difference between the two false alarm rate measures (p=0.249, Wilcoxon signed-rank test, n=31).

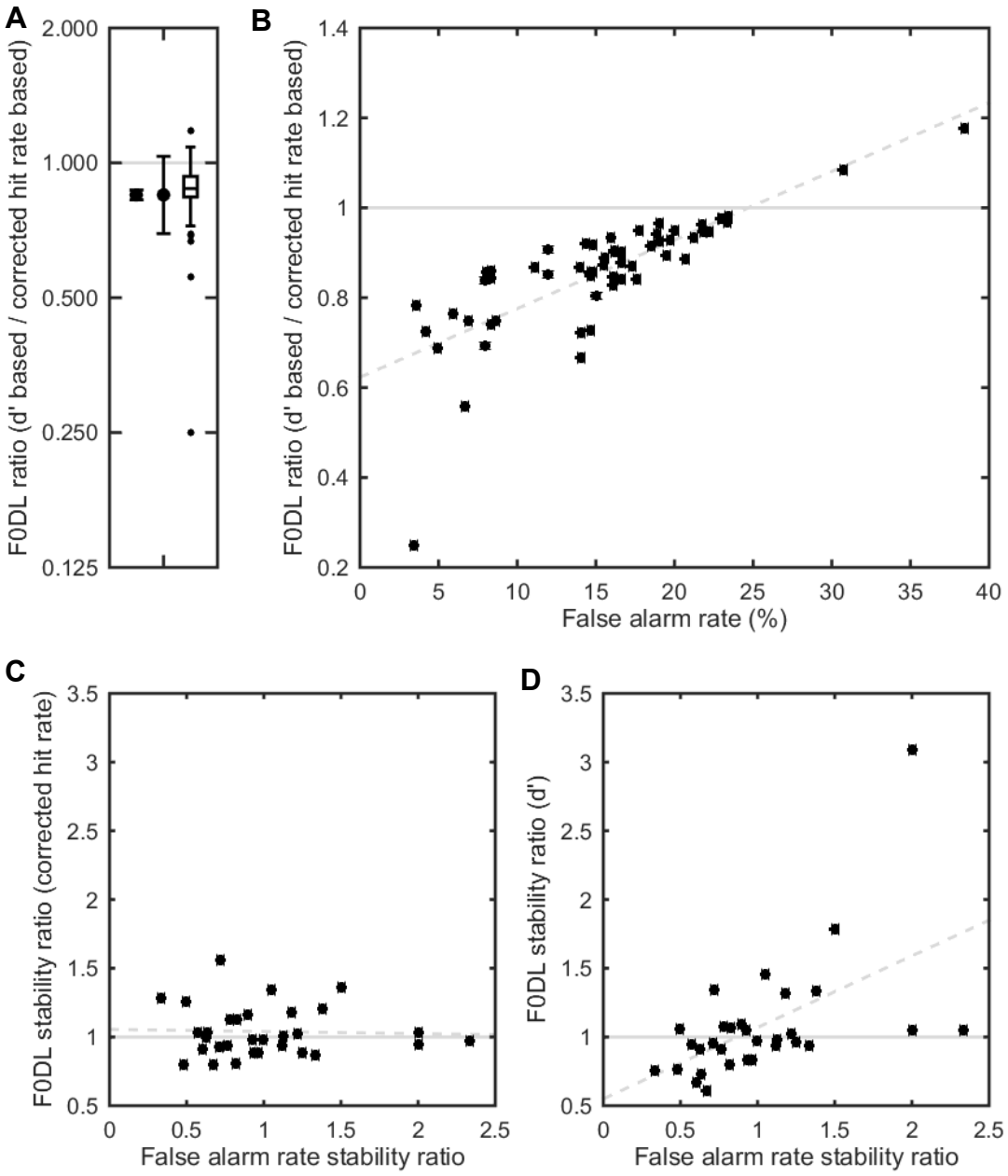


Figure 4.3 Comparison of F0DL calculations

Two different F0DL calculations were compared. (A) The ratio between the F0DL based on d' and the F0DL based on corrected hit rate of the same measure. The left

error bar indicates the mean of the ratio (0.85) and SEM (1.03). The middle error bar indicates the mean and SD (1.22), with the box plot to the right. F0DLs based on d' are significantly smaller from F0DLs based on corrected hit rate (p=6.4e-11, Wilcoxon signed rank test, one side, n=62). (B) The ratio between two different F0DL calculations, plotted against the false alarm rate of the same measure. A linear regression model shows the fitted ratio as 1.526 · (False alarm rate) + 0.623 (R^2=0.611, adjusted R^2=0.604, and p=6.7e-14 for the linear coefficient), as indicated by the grey dashed line (same for C and D). The model shows the variation of the ratio is largely dependent on the false alarm rate. And when the false alarm rate is around 24.7%, the ratio is around 1, where the two F0DL calculations give the same estimation. (C) Stability ratio of corrected hit rate based F0DL plotted against stability ratio of false alarm rate. Corrected hit rate based F0DL does not change significantly while false alarm rate changes (linear regression model, F0DL ratio = -0.016 · (False alarm rate ratio) + 1.055, R^2=0.0016, adjusted R^2=-0.0329, and p=0.832 for the linear coefficient). (D) Stability ratio of d' based F0DL plotted against stability ratio of false alarm rate. The d' based F0DL increases significantly while false alarm rate increase (linear regression model, F0DL ratio = 0.5202 · (False alarm rate ratio) + 0.5492, R^2=0.2953, adjusted R^2=0.2710, and p=0.0016 for the linear coefficient).

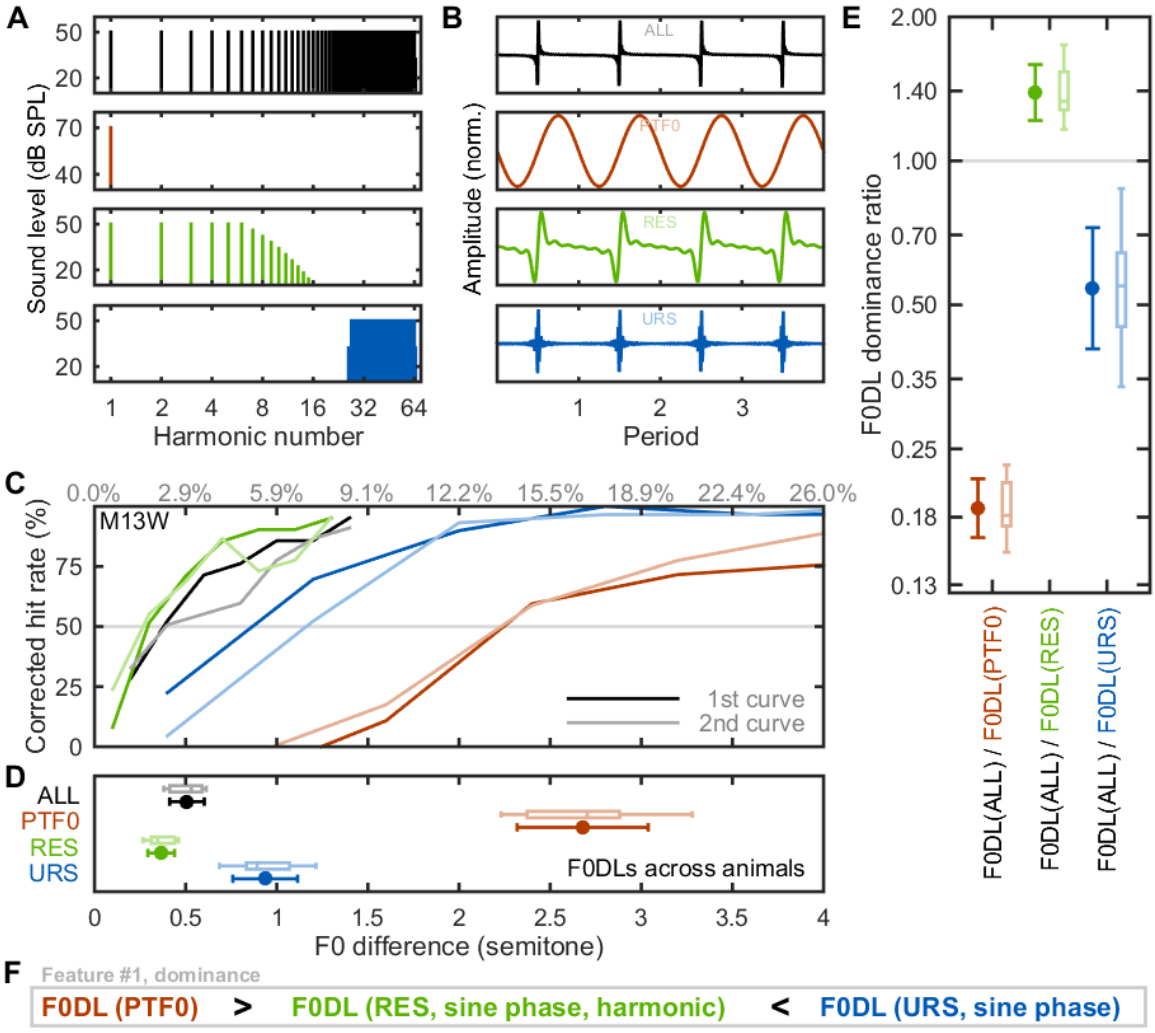


Figure 4.4 RES dominate marmoset pitch strength, similar to humans (corrected hit rate based F0DL calculation)

Spectra (A) and waveforms (B) of the background sounds used in marmoset F0DL measurements, for ALL (black), PTF0 (red), RES (green), and URS (blue, noise masker not shown). (C) Example psychometric curves from the subject M13W under the 4 conditions in (A). Darker and lighter lines indicate the 1st and the 2nd measured curves,

respectively. The grey line indicates 50% corrected hit rate. (D) F0DLs under each condition across all tested animals and measurements. Error-bars indicate the mean values and SDs, with box plots above. (n=8, for each) (E) F0DL dominance ratios, defined as the ratio between F0DL of all harmonics presented together and F0DL measured under each decomposed condition (n=8, for each). The grey line shows a reference ratio equal to 1. The error-bars indicate the mean values and SDs, with box plots on the right. (n=8, for each). F0DLs under PTF0 condition are significantly higher than F0DL of ALL ($p=0.0039$, n=8). F0DLs under URS condition are significantly higher than F0DL of ALL ($p=0.0039$, n=8). F0DLs under RES condition are significantly lower than F0DL of ALL ($p=0.0039$, n=8). (F) The summary of human-like primary feature #1: lower resolved harmonics (RES) have a stronger pitch strength, thus a smaller F0 discrimination threshold, compared to both pure tones at the fundamental frequency (F0) and higher unresolved harmonics (URS).

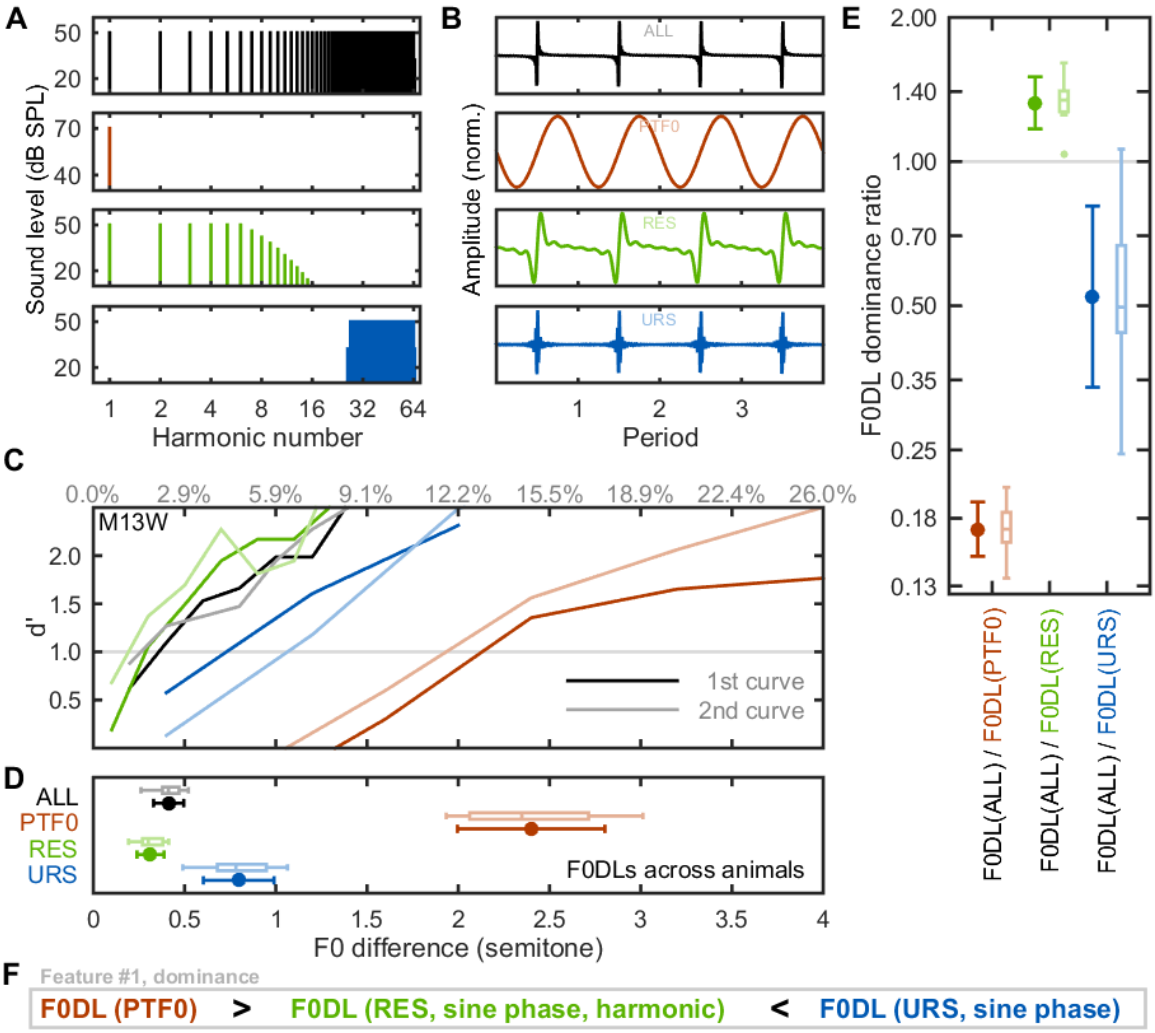


Figure 4.5 RES dominate marmoset pitch strength, similar to humans (d' based F0DL calculation)

Spectra (A) and waveforms (B) of the background sounds used in marmoset F0DL measurements, for ALL (black), PTF0 (red), RES (green), and URS (blue, noise masker not shown). (C) Example psychometric curves from the subject M13W under the 4 conditions in (A). Darker and lighter lines indicate the 1st and the 2nd measured curves, respectively. The grey line indicates a reference d' equal to 1. (D) F0DLs under each

condition across all tested animals and measurements. Error-bars indicate the mean values and SDs, with box plots above. (n=8, for each) (E) F0DL dominance ratios, defined as the ratio between F0DL of all harmonics presented together and F0DL measured under each decomposed condition (n=8, for each). The grey line shows a reference ratio equal to 1. The error-bars indicate the mean values and SDs, with box plots on the right. (n=8, for each). F0DLs under PTF0 condition are significantly higher than F0DL of ALL ($p=0.0039$, n=8). F0DLs under URS condition are significantly higher than F0DL of ALL ($p=0.0078$, n=8). F0DLs under RES condition are significantly lower than F0DL of ALL ($p=0.0039$, n=8). (F) The summary of human-like primary feature #1: lower resolved harmonics (RES) have a stronger pitch strength, thus a smaller F0 discrimination threshold, compared to both pure tones at the fundamental frequency (F0) and higher unresolved harmonics (URS).

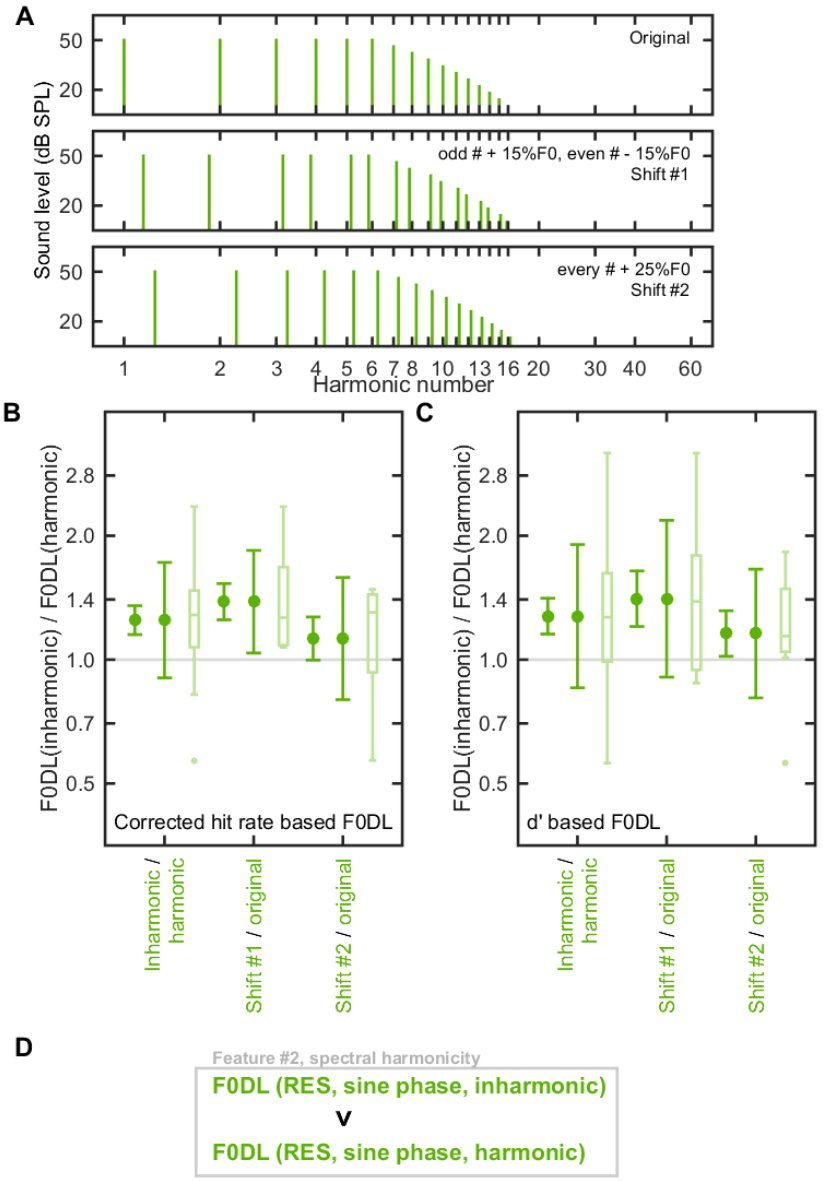


Figure 4.6 F0DLs of RES are sensitive to the quality of spectral harmonicity in marmosets, similar to humans

(A) Spectra of the background sounds used for testing harmonicity sensitivity on RES. Vertical ticks serve as references for integer harmonic numbers. (B, C) F0DL

harmonicity ratios, defined as ratios between inharmonic F0DL and harmonic F0DL. The grey line shows a reference ratio equal to 1. The error-bars indicate the mean values with SDs, with box plots on the right. (n=16, for inharmonic/harmonic, n=8, for shift#1/original, and shift#2/original). Both ratios of corrected rate based F0DL and ratios of d' based F0DL are shown (D) The summary of human-like primary feature #2: pitch perception based on resolved harmonics is sensitive to the fidelity of spectral harmonicity.

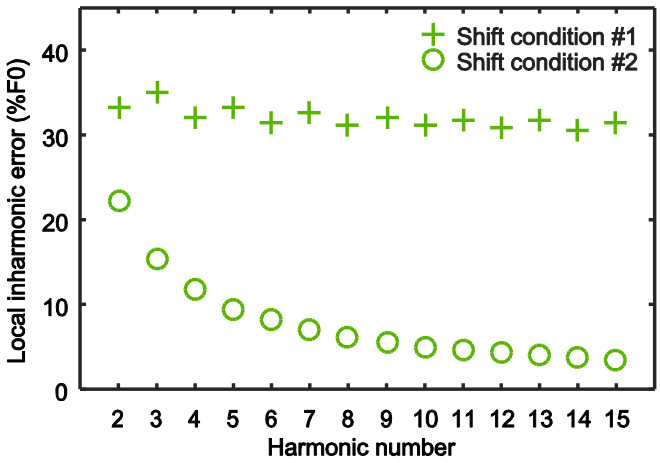


Figure 4.7 Inharmonicity produced by shift conditions

For each of the inharmonic shift conditions in the present experiment, we measured the degree of deviation from harmonicity around each component. To do this, for each component together with its neighboring two components, we found the nearest three harmonic template that best approximated the inharmonic complex. The frequency difference between each of the three inharmonic components and its corresponding harmonic component was calculated as individual inharmonic error. A local inharmonic error measure was calculated by summing the individual inharmonic errors across all three adjacent components and dividing by the F0 of the harmonic template. As shown in the figure, shift condition #1 introduces consistently significant amounts of inharmonicity across the entire RES frequency range. However, shift condition #2 can only introduce significant amounts of inharmonicity for the first several harmonics, and the local inharmonic error drops below 10% of F0 after the 5th harmonic. In humans, the spectral

region that contributes most to the globally assembled pitch perception is around 1st-4th harmonic [Plack et al 2005]. However, in marmosets, the spectral region that contributes most to the globally assembled pitch might be composed of the higher numbered harmonics, although still in the RES range. If that is true, then, for marmosets, Shift condition #2 approximates a harmonic series inside this region, while Shift condition #1 remains inharmonic across the entire RES range. Our data show that shift condition #1 introduced a significant increase in F0DL while shift condition #2 only shows an increase trend, and may suggest that the spectral region that contributes most to the globally assembled pitch in marmosets might be higher than those in humans, at least for an F0 of 440 Hz.

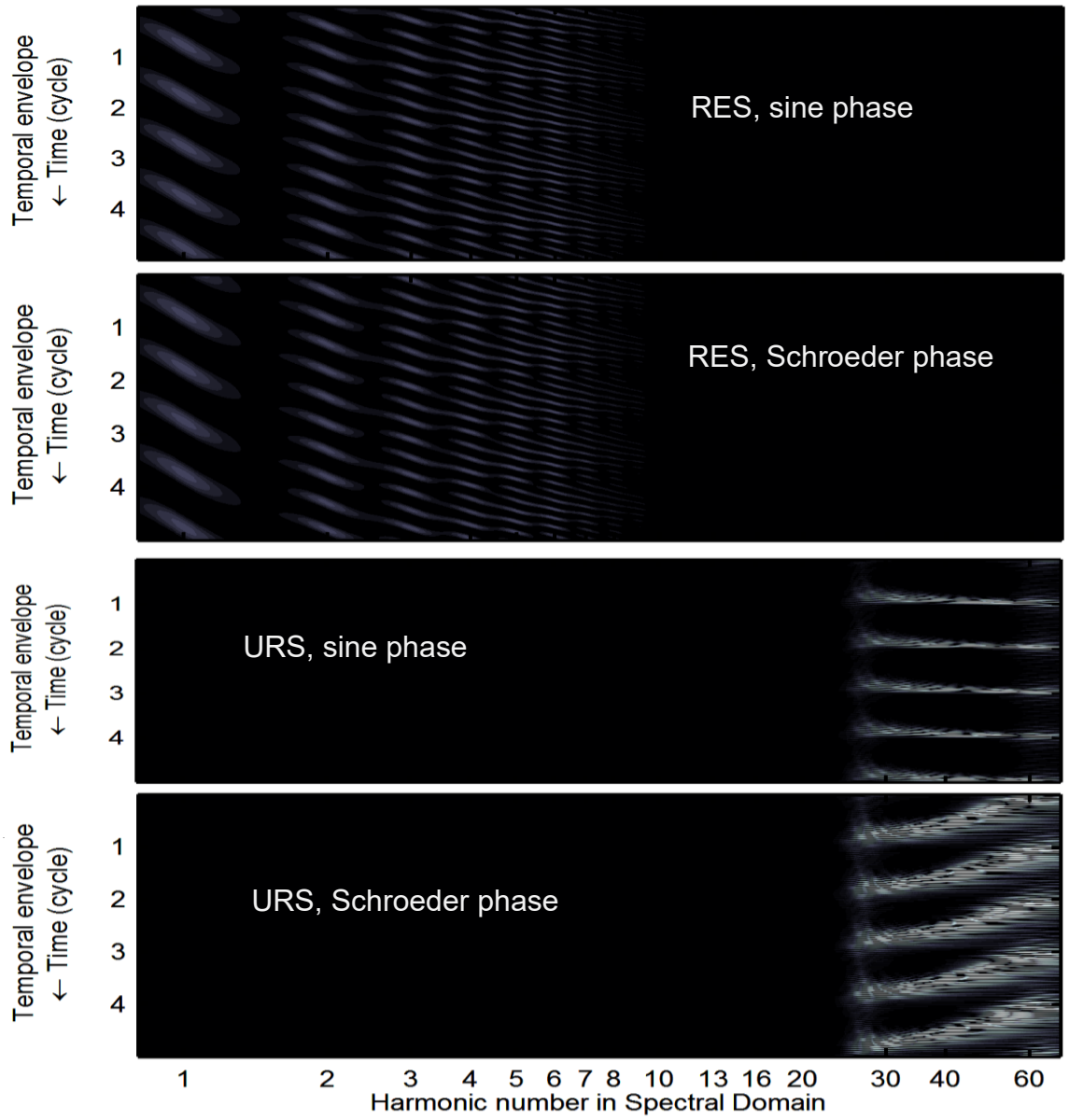


Figure 4.8 Cochleograms of stimuli with Schoeder or sine phase

The spatiotemporal activity pattern of marmoset auditory peripheries. Five F0 cycles were shown along the vertical axis, against harmonic numbers along the horizontal axis. Schroeder phased RESs do not alter general temporal envelope structures on individual auditory channel, similar to their sine phased counterparts. Schroeder phased

URSs introduce a smear on temporal envelope cue of periodicity, whereas their sine phased counterparts keep shape temporal peaks within each temporal cycle.

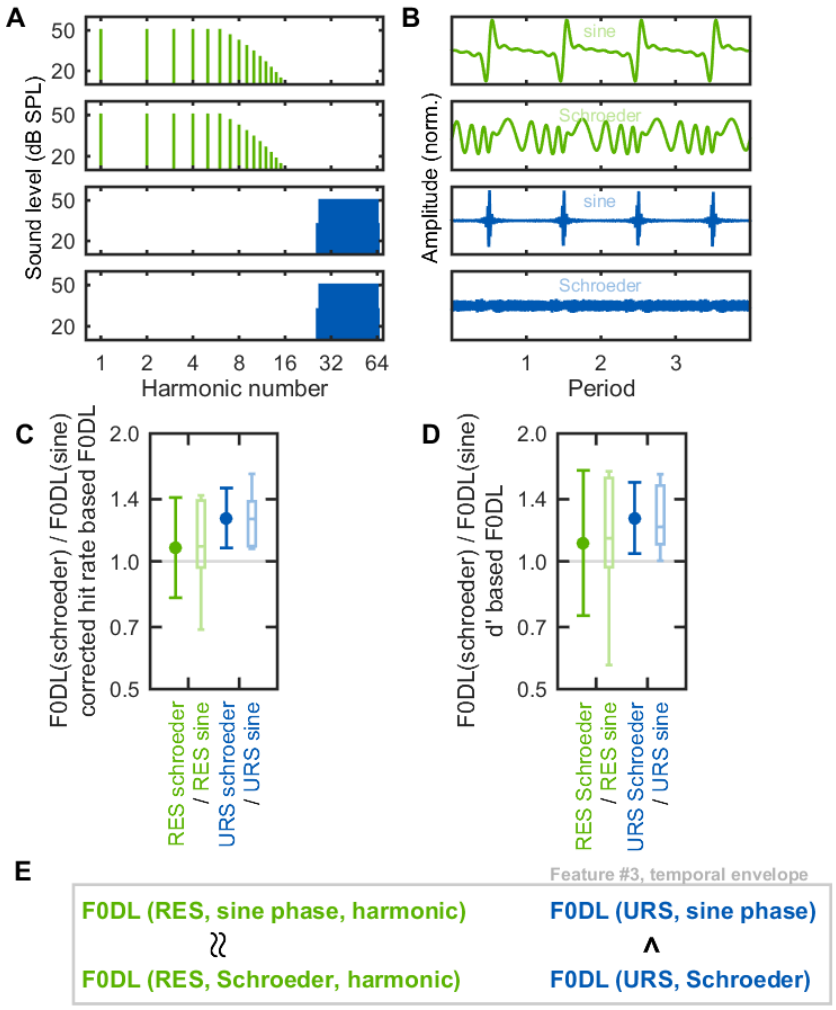


Figure 4.9 F0DLs of URS are sensitive to the salience of temporal envelope cues in marmosets, similar to humans

Spectra (A) and waveforms (B) of the background sounds used for testing URS sensitivity to the salience of temporal envelope cues. (C, D) F0DL Schroeder/sine ratios, defined as ratios between F0DL measured using Schroeder phase sounds and F0DL measured using sine phase sounds. The grey line shows a reference ratio equal to 1. The error-bars indicates the mean values with SD, with box plots on the right (n=6, for each).

(C) Ratios calculated on corrected hit rate based F0DL. (D) Ratios calculated on d' based F0DL. (E) The summary of human-like primary feature #3: pitch perception based on unresolved harmonics is sensitive to the saliency of temporal envelope cues.

Subject / Order	M62U	M13W	M11X	M4Y
1	PTF0	PTF0	PTF0	PTF0
2	ALL	ALL	ALL	ALL
3	URS, sine phase	RES, shift #2	URS, sine phase	RES, sine phase
4	RES, shift #1	RES, sine phase	URS, Schroeder	RES, shift #2
5	RES, shift #2	RES, Schroeder	RES, sine phase	RES, Schroeder phase
6	RES, sine phase	RES, shift #1	RES, Schroeder	RES, shift #1
7	URS, Schroeder	URS, Schroeder	RES, shift #2	URS, sine phase
8		URS, sine phase	RES, shift #1	URS, Schroeder

Table 4.1. Testing order of different conditions on each animal.

	Subject	M4Y		M11X		M62U		M13W		mean (arithmetic)	mean (geometric)
	Measure #	1st	2nd	1st	2nd	1st	2nd	1st	2nd		
F0DL (semitone, corrected hit rate based)	ALL	0.500	0.433	0.562	0.579	0.600	0.613	0.380	0.392	0.51	0.50
	F0	2.841	2.850	2.567	2.504	2.911	3.280	2.244	2.230	2.68	2.66
	RES	0.430	0.344	0.350	0.331	0.461	0.450	0.293	0.267	0.37	0.36
	URS	0.904	0.800	1.215	0.978	0.686	0.880	0.867	1.162	0.94	0.92
	RES_sft_#1	0.461	0.543	0.383	0.355	0.542	0.612	0.523	0.628	0.51	0.50
	RES_sft_#2	0.244	0.283	0.369	0.464	0.683	0.545	0.414	0.390	0.42	0.40
	URS_Sch	0.981	0.922	1.300	1.343	1.200	1.067	1.200	1.867	1.23	1.21
	RES_Sch	0.415	0.407	0.242	0.329			0.418	0.371	0.36	0.36
F0DL (semitone, d' based)	ALL	0.428	0.403	0.420	0.399	0.511	0.521	0.359	0.261	0.41	0.40
	F0	2.615	2.567	2.012	2.113	2.813	3.012	2.131	1.934	2.40	2.37
	RES	0.413	0.251	0.294	0.309	0.371	0.390	0.287	0.193	0.31	0.31
	URS	0.855	0.709	1.042	0.833	0.651	0.490	0.730	1.064	0.80	0.77
	RES_sft_#1	0.381	0.503	0.284	0.271	0.511	0.543	0.460	0.613	0.45	0.43
	RES_sft_#2	0.232	0.253	0.318	0.336	0.610	0.467	0.386	0.352	0.37	0.35
	URS_Sch	0.857	0.801	0.325	1.004	1.302	1.255	1.170	1.572	1.04	0.95
	RES_Sch	0.404	0.394	0.167	0.299			0.378	0.314	0.33	0.31
False alarm rate (%)	ALL	14.8%	19.8%	8.6%	4.9%	12.0%	14.7%	22.2%	14.1%	0.14	0.13
	F0	14.4%	16.2%	3.6%	8.3%	19.1%	14.8%	17.8%	11.1%	0.13	0.12
	RES	21.8%	14.7%	8.0%	16.0%	15.1%	14.0%	23.5%	14.1%	0.16	0.15
	URS	22.2%	20.7%	14.7%	12.0%	20.0%	6.7%	17.6%	18.5%	0.17	0.16
	RES_sft_#1	16.1%	19.1%	8.3%	6.0%	18.9%	15.6%	16.7%	23.0%	0.15	0.14
	RES_sft_#2	21.8%	19.5%	8.3%	4.2%	16.7%	8.1%	21.2%	16.2%	0.14	0.13
	URS_Sch	15.5%	17.3%	3.5%	6.9%	30.8%	38.5%	23.1%	16.7%	0.19	0.15
	RES_Sch	23.3%	23.3%	8.0%	12.0%			16.7%	16.1%	0.17	0.16

Table 4.2. Thresholds and false alarm rates from all measures

All F0DL measurements, listed in both corrected hit rate based F0DLs and d' based F0DLs, as well as false alarm rates of all measurements. Mean value of F0DLs and false alarm rates are provided in both arithmetic mean and geometric mean.

5. PITCH: FROM PERCEPTION TO PHYSIOLOGY

5.1. Summary

In this chapter, several topics beyond pitch perception mechanisms are discussed. First, the development of a silent two-photon imaging approach in awake marmosets is an ongoing effort aiming to see more details from the previously discovered marmoset cortical pitch center. Second, inactivation of this pitch center during pitch discrimination task may reveal whether this pitch center plays a causal role in marmoset pitch perception. Furthermore, testing of higher-order perceptions based on but beyond basic pitch perception is also discussed. These topics may serve as the future directions for the next generation of pitch and music element related studies in marmosets.

5.2. Introduction

To understand the neural mechanisms underlying human pitch perception, the most direct approach would be recording single neurons in human brains. However, in practice, this requires invasive procedures and the chance to do so is highly limited. Although, noninvasive functional imaging can record indirect populational signals based on metabolic changes, the nature and the quality of the signal are not even nearly similar to single unit recordings. An appropriate animal model thus may provide extensive experimental accessibility to record single neurons, test hypotheses, and provide further clues for neural mechanisms for pitch in humans.

In previous chapters, behavioral evidence has been shown that the marmoset monkey, comparing to other previously tested non-human mammals, has more human-like pitch perception features. These features include: (1) marmoset monkeys are sensitive to pitch change from missing fundamental complex sounds at only one semitone pitch difference; (2) the pitch strength of complex sounds in marmosets is dominated by resolved harmonics; (3) the pitch of resolved harmonics in marmosets is sensitive to the fidelity of spectral harmonicity. These findings may put marmosets as a more appropriate nonhuman animal model to study human-like pitch perception, as illustrated in figure 5.1.

In addition to these perceptual data, single unit recordings in marmoset auditory cortex has revealed a cortical pitch center contains neurons responsive to pitch-evoking sounds in humans [Bendor and Wang 2005]. Depending on F0 and harmonic resolvability, these neurons extract pitch using either spectral harmonicity cue or temporal envelope cue [Bendor et al 2012]. In humans, an apparently homologous cortical pitch processing center has been localized by fMRI, at the same anterolateral low-frequency border of primary auditory cortex [Norman-Haignere et al 2013, Penagos et al 2004].

These datasets have led to several interesting questions. One of them is whether there's a functional microarchitecture or laminar distribution within the very small cortical pitch center ($<1mm^2$). To answer this question, single unit recording may not provide enough recording density and spatial certainty. I have been working on developing a silent two-photon imaging approach aiming to record a large density of neurons simultaneously and to aim investigating this question. This topic is discussed in chapter 5.3.

Another question is whether the cortical pitch center plays a causal role in marmoset pitch perception, as linking the lower-left block to the lower-right block in figure 5.1. This topic is discussed in chapter 5.4.

Furthermore, our current behavior dataset only covers the condition of perceiving a single pitch around F0=440Hz. Pitch perception of other F0s and perceptions beyond a single pitch, like octave generalization, consonance, and dissonance perception, are discussed in chapter 5.5.

5.3. Imaging Neuronal Functions in Marmoset Cortical Pitch Center with a Silent Two-photon Microscope

5.3.1 Development of a silent two-photon microscope

Long-term chronic two-photon imaging has been deployed in rodents to study functional neuronal circuits in auditory cortex [Rothchild et al 2010, Bandyopadhyay et al 2010, Chen et al 2011, Letzkus et al 2011, Bathellier et al 2012, Rothchild et al 2013, Issa et al 2014, Barnstedt et al 2015, Kato et al 2015, Maor et al 2016, Deneux et al 2016]. However, two-photon imaging setup is usually quite loud by nature. And to the best of my knowledge, there has been no two-photon microscope that can be considered fully “noise-free” or “silent”. To apply the two-photon imaging technique to study auditory cortex without acoustic artifacts, a quiet imaging setup is required. The main noise sources of a two-photon imaging setup are illustrated in figure 5.2 and discussed in the following paragraphs.

A standard implementation of two-photon microscopy utilizes a pair of mechanically oscillating mirrors to scan the laser beam [Grewe and Helmchen 2009] and

thus introduces considerable acoustic noises on distinct frequencies. And when a high scanning speed is desired, a resonant scanner is usually chosen. The resonant frequency of a commercially available resonant scanner is usually around 7-8kHz (illustrated in figure 5.2 as the black arrow), which, unfortunately, is within the marmoset dominant vocalization F0 range [Agamaite et al 2015]. In addition, marmosets are very sensitive to such a frequency [Osmanski and Wang 2011, Osmanski and Song et al 2016]. Instead of using a mechanical resonant scanner in our setup, a pair of acousto-optical deflectors (AOD) were chosen to deflect the laser beam into different angles. An AOD scanner does not move anything mechanically during scanning and thus generates no audible mechanical noise.

Besides scanning noise which is mainly composed of narrowband sounds at distinct frequency peaks, power supply and cooling system of a Ti:Sapphire laser system generate significant broadband noises at low and middle frequencies. Such broadband sounds are typically quantified as power density (in dB SPL/Hz) and can only be compared to a species' audiogram, which is based on the power in a single-frequency tone (in dB SPL), by assuming some resolving bandwidth for stimulus intensity. Therefore, to examine whether a broadband sound is audible to marmosets, the spectrum of the sound is weighted by marmoset auditory peripheral tuning bandwidth (ERB) [Osmanski et al 2013] and then compared to marmoset audiogram [Osmanski and Wang 2011]. Noises were measured one meter away from both the laser power supply and cooling system (Coherent, CA, Chameleon Vision S) using a handheld sound level meter (Brüel & Kjaer, 2250 with microphone 8190) with an amplified analog output to a data acquisition card (National Instruments, NI PCIe-6323). The measuring system's weighted

noise floor was calibrated as at least 3 dB lower than marmoset audiogram across the entire spectrum. The weighted cooling and power supply noise is illustrated as the green area in figure 5.2, and was consistently around 20 dB higher than marmoset audiogram at low and middle frequencies up to ~10kHz. To get rid of this cooling and power supply noise, we isolated the cooler and the power supply of the laser outside a double-wall acoustic chamber (IAC industries) and connected the cooling and connection lines through U-shaped tubes into the chamber to minimize noise leaks.

After isolating the laser box from the cooler and power supply, a high-frequency sound from the laser box became noticeable during laser operation, illustrated as the orange area in figure 5.2. This noise was due to piezo stages operating inside the laser cavity to maintain maximal power output of the laser. It had two major spectral peaks. One was around 6-7 kHz, and the other was around 13 kHz. A sound enclosure was designed with polyoxymethylene plastic boards layered with sound-absorbers to cover the laser box and isolate the noise.

After these efforts. Our setup worked below marmoset audiogram across the entire spectrum, illustrated as the green area in figure 5.2. Since auditory filters are wider at louder sound levels, and ERBs were measured at moderate sound levels, the weighted sound level estimation probably overestimates the noise level when it is actually at a level near to the audiogram. Thus, our setup works conservatively below the marmoset audiogram and can be considered as noise-free or silent.

Calcium indicators used for chronic two-photon imaging are typically genetically encoded (e.g. GCaMP6). They are inherently nonlinear, and report sustained firings more robustly than brief firings [Chen et al 2013]. Interestingly, sustained neuronal firing

during an acoustic stimulus (as opposed to an onset response) is only seen in awake marmosets in auditory cortex [Wang et al 2005]. To maximize the chance for robust signal detections and interpretations in marmoset auditory cortex, it is necessary to keep experimental animal awake. However, imaging under awake condition may introduce extensive motion artifacts. Single pair AOD based scanning is typically designed to do two-dimensional (2D) random access scanning [Grewe et al 2010], which is very vulnerable to motion artifacts and cannot guarantee selected points are always within desired structures during the motion. We extended our AOD scanning modes from 2D random access scanning to 2D raster scanning at the video rate, and to 3D multi-layer raster scanning. By recording the full frame in the field of view (raster scanning) at a fast speed, the chance of realigning frames back to each other to correct motion artifacts was maximized.

To sum up, I designed and built a silent two-photon imaging system without any audible noises to marmosets, while maintaining advanced scanning speed and flexibility. We thus call our system "Flexible, Agile, and Noise-free Two-photon AOD Scanning Imaging in Awake animals" (FANTASIA). This approach can also be generalized to image auditory cortex in other species without acoustic artifacts.

5.3.2 Surgical design

Long-term chronic two-photon imaging has been widely applied in rodents to study functional neuronal circuits underlying behaviors [Peron et al. 2015]. However, the application in non-human primates has been very limited [O'Shea et al 2016]. Among

primates, marmosets are highly social animals and have a flat cortical surface that is ideally suited for optical imaging studies [Miller et al 2016].

Chronic artificial dura based optical window has been developed for intrinsic imaging in non-human primates [Roe 2007]. The feasibility of penetrations by glass pipettes or metal electrodes has also allowed artificial dura to be applied in fields such as primate optogenetics [Ruiz et al 2013]. However, since the customized silicone artificial dura has the thickness generally around 200 microns, and is difficult to be made optically flat, a standard artificial dura based optical window may introduce too much optical aberration and thus contaminate functional signals when cellular resolution imaging is desired. To reduce the optical aberration introduced from the artificial dura based optical window, we customized our mold to reduce the thickness of silicone based artificial dura from ~ 200 um to ~60um. Furthermore, we designed a mold utilizing tungsten carbide clamps from a micrometer to maximize the flatness of the artificial dura and further reduced optical aberration.

To start preparing an optical window for two-photon imaging, a small craniotomy (~8mm in diameter) was made above the auditory cortex while the animal was anesthetized. The location of the primary auditory cortex (A1) was confirmed beforehand by neural recording with tungsten electrodes through miniature holes (~1mm in diameter) and the intact dura using the standard techniques in our lab [Lu et al 2001]. A durotomy was made over the targeted cortical region containing the primary auditory cortex. Afterward, a silicone-based pre-molded artificial dura was implanted, secured, and sealed to the craniotomy edge by silastic (Kwik-sil, WPI inc).

Glass pipettes can penetrate the artificial dura and inject adeno-associated virus (AAV) carrying calcium sensor like GCaMP6 [Chen et al 2013] into the targeted cortical tissue. Before two-photon imaging sessions take place, a #0 coverslip was placed on top of the artificial dura to reduce the movement artifacts in optical recording sessions.

5.3.3 Preliminary results

We have conducted preliminary experiments to demonstrate that we can maintain the artificial dura in awake marmosets for more than 70 days (figure 5.3). This artificial dura based optical window is chronic, removable, easy to maintain, and suitable for high-resolution two-photon imaging in awake marmosets. we injected adeno-associated virus carrying GCaMP6s and eGFP into the auditory cortex of a marmoset monkey. The testing injection sites are shown in figure 5.4. through wide field imaging. GFP expression can be seen weeks after the initial injections.

Around the GCaMP6s injection site, we performed two-photon imaging at video rate through the artificial dura window under awake condition. Clear cellular structures could be resolved. Motion artifacts could be removed by aligning frames to an averaged template. For an exemplary labeled cell soma shown in figure 5.5, when playing the same sound sequence for twice, the fluorescence traces were generally repeatable. We suspected this exemplary cell is an inhibitory interneuron with high spontaneous firing rate.

It has been suggested when using at a low titer, synapsin promoter has a trend to label inhibitory cells sparsely, but not pyramidal cells densely in upper layers of the cortex [Nathanson et al 2009]. This is also consistent with recent virus labeling results in

primates [Watakabe et al 2015, Sadakane et al 2015, Seidemann et al 2016]. Alternatively, CaMKII promoter may have a trend to label more upper layer excitatory cells in both marmosets and macaques [Watakabe et al 2015, Seidemann et al 2016]. We plan to test more AAV serotypes (including serotype DJ [Grimm et al 2008]) with CaMKII promoter and try to enhance the labeling efficiency and expression level in marmoset cortex. Moreover, another possibility is to use double-transfection strategy to separately control expression level and expression specificity [Sadakane et al 2015]. We also plan to test double-transfection strategies like AAV-CaMKII-cre + AAV(DJ)-EF1a-flex-GCaMP6s to enhance both expression level and expression pattern in marmosets.

To sum up, we have developed a silent two-photon microscope, and a chronic, removable, artificial dura based optical window that is suitable and optimized for high-resolution chronic two-photon imaging in awake marmoset monkeys. Our preliminary results showed that we can record single neurons functionally in auditory cortex, although the viral labeling efficiency needs to be further optimized. Such a technical approach may open the possibilities to investigate neuronal circuits underlying higher-order auditory behaviors in awake marmosets, such as pitch perception.

5.3.4 Potential hypotheses

Single unit recordings in marmosets have shown neurons in marmoset cortical pitch center use both temporal and spectral information to extract pitch. It was also suggested that pitch information is extracted either using temporal information for low pitch sounds composed of high order harmonics or using spectral information for high pitch sounds with low order harmonics [Bendor et al 2012]. Furthermore, temporal

periodicity tuned neurons tend to cluster within pitch center [Bendor and Wang 2010], whereas neurons respond more to spectral harmonic templates than to single frequency components were not restricted to pitch center [Feng 2013].

One interesting question is whether there is any functional microarchitecture within the pitch center. Single unit recordings may not provide enough spatial certainty to answer this question. Alternatively, two-photon imaging may record densely packed neurons with spatial certainty once labeling efficiency is optimized. The technique thus may open an opportunity to search for any functional microarchitecture. One hypothesis is that the importance of spectral information may gradually decrease and the importance of temporal information may gradually increase when the recording location moves towards the low-frequency side on the tonotopic axis.

Another interesting question is whether there is a functional laminar difference in pitch processing. In marmoset auditory cortex, phase locking to temporal periodicity tends to be limited to very low periodicities that is below the classical pitch range [Lu et al 2001]. Precise temporal pitch information is most likely to be converted into a rate code subcortically and relayed to the input middle layer (layer 4) of the cortex through thalamo-cortical connections. For spectral information, since harmonic template neurons were found mostly in upper layers (layer 2 and 3) outside pitch center [Feng 2013], they may have interconnections with pitch neurons within the pitch center. It is possible that the temporal pitch arrives in layer 4 of pitch center through thalamo-cortical connections, whereas the spectral pitch is finally processed by layer 2 and 3 neurons in pitch center, potentially benefiting from intercortical connections with harmonic template neurons

outside the pitch center. Recording neurons from different layers may give a clue to test this hypothesis.

5.4. Is Cortical Pitch Center Necessary for Pitch Perception?

In previous chapters, we have provided behavioral evidence that marmoset monkeys possess human-like pitch perception mechanisms, thus established a similarity link between human pitch perception and marmoset pitch perception (illustrated in figure 5.1). However, to use data recorded from marmoset pitch neurons to explain human pitch perception, a question needs to be answered first is whether the cortical pitch center plays a causal role in marmoset pitch perception (illustrated as the dashed line arrow in figure 5.1). It is interesting to inactivate the pitch center while having the subject doing a pitch task to see whether the animal's performance is altered or not. The inactivation can be optical, thermal, pharmacological, or surgical. Each may have its own requirements on behavior task design.

5.5. Marmoset Pitch and Related Perceptions Beyond Current Dataset

5.5.1 Pitch mechanisms' dependence on fundamental frequency

Neural recording data from pitch neurons in marmosets have suggested that there is an F0 dependence boundary around ~450 Hz. For F0s lower than this boundary, pitch neurons rely on temporal cues more, and for F0s higher than this boundary, pitch neurons rely on spectral cues more [Bendor et al 2012]. The current behavior dataset suggests that, around this boundary F0, for an F0 = 440 Hz, both types of cues exist. For spectral cues on resolved harmonics, F0DL is roughly around half a semitone. For temporal cues

on unresolved harmonics, F0DL is roughly around one semitone. Currently, we are also collecting F0DL data from more subjects under all harmonics conditions for more F0s. Our preliminary data suggest that, for an F0=110Hz, F0DL is 1.32 semitones (n=3), for an F0=220Hz, F0DL is 1.03 semitones (n=3), for an F0=880Hz, F0DL is 0.596 semitones (n=3). It is interesting to investigate further and test (1) whether the octave between 220 Hz and 440 Hz is a transition point that spectral pitch becomes available (2) whether F0DL based on temporal cues are generally around one semitone and F0DL based on spectral cues are generally around half a semitone.

5.5.2 Music element related perceptions beyond a single pitch

The current study focused on how a single pitch is processed from a complex sound. In the real world, especially in music, the relationship between or among pitches can evoke further interesting perceptions, like octave generalization, consonance, or dissonance [McDermott and Oxenham 2008]. Such topics in non-human animals are generally still open questions. It has been suggested songbirds (European starling) use spectral shape, not pitch, for sound pattern recognition [Bregman et al 2016]. Inside the study, the sound pattern was a sequence of pitch notes in two semitone steps and with different spectral shapes. However, it is unclear whether the species can discriminate two semitone pitch change easily in the first place. It is interest to have a control experiment to first show that whether the species can perceive the pitch change in such a step first. Thus, out current dataset may provide a basis and a guidance for further music-related experimental design in marmosets. As the octave above 440Hz may have a precision to allow Western musical melody discrimination.

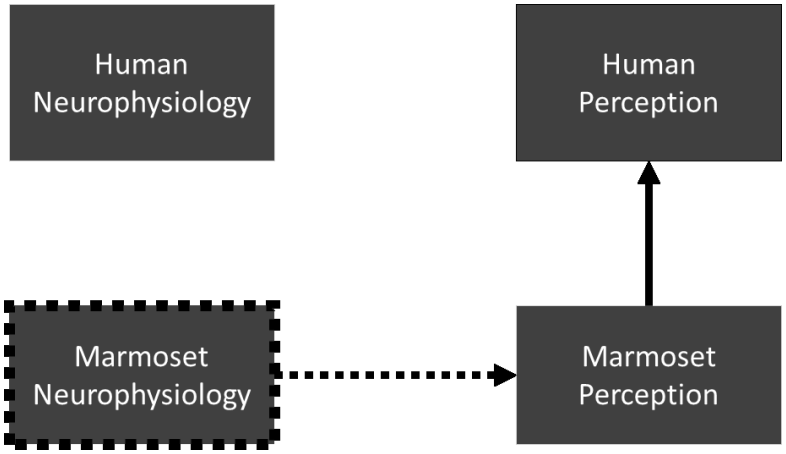


Figure 5.1 The summary of pitch studies

The current study suggests marmoset also possess human-like pitch perception mechanisms, and thus provides a similarity link (solid arrow) between human pitch perception (upper right block) and marmoset pitch perception (lower right block). Recording human auditory cortex (upper left block) extensively using invasive neurophysiological methods is not quite feasible. To use neurophysiology data recorded from marmoset pitch center (lower left block) to explain human pitch perception (upper right block), it is necessary to show a causal link (dashed arrow) between marmoset pitch center (lower left block) and marmoset pitch perception (lower right block). To record more details within the small marmoset cortical pitch center, a novel technique beyond single unit recording is required, illustrated as the dashed box.

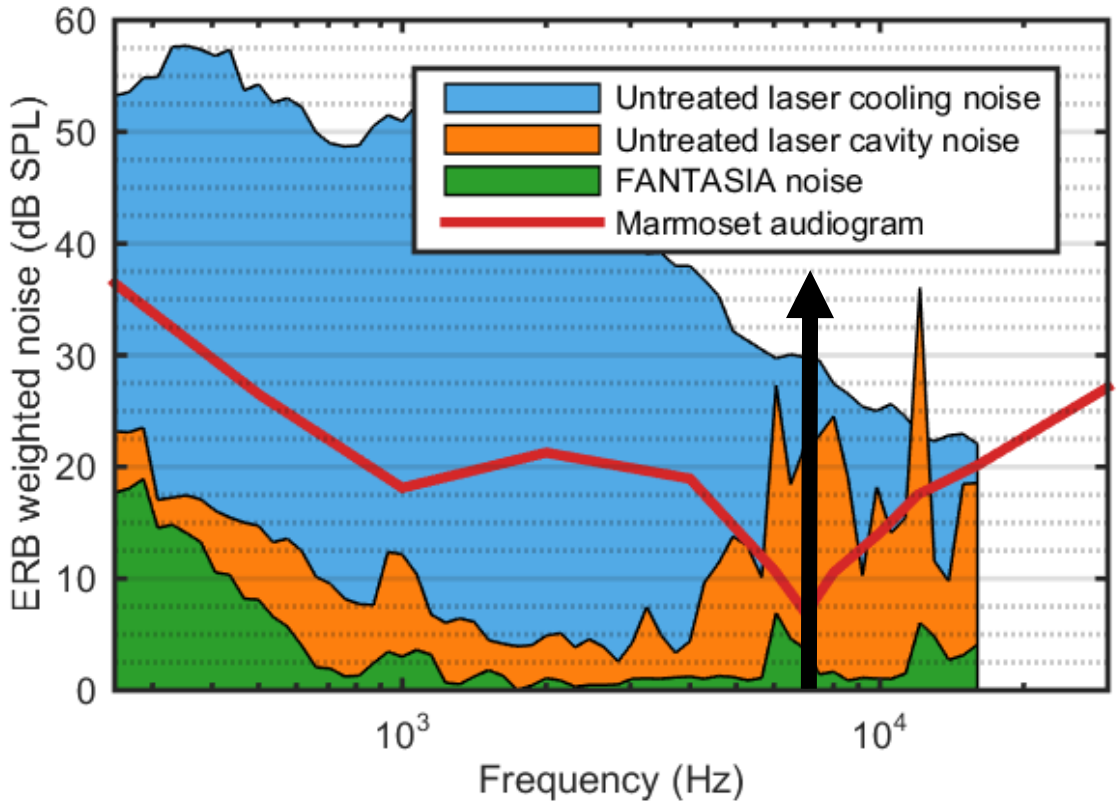
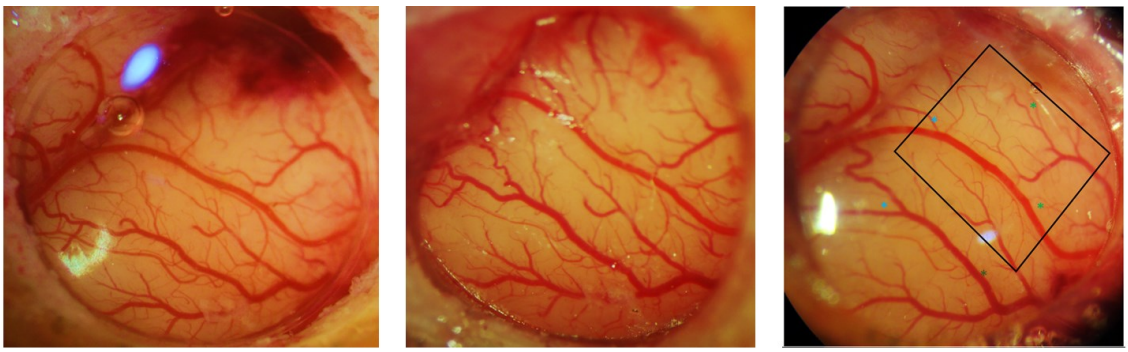


Figure 5.2 Acoustic noise floor of our FANTASIA microscope

The major noise sources of a standard fast two-photon microscope are illustrated with their spectral signatures: (1) mechanically resonating scanning mirror (black arrow); (2) operating piezo stages inside the laser cavity (orange area); (3) cooling and power supply system of the ultrafast laser (blue area). Marmoset audiogram (red line) is plotted on top of the auditory-tuning-bandwidth weighted noises for direct comparison. Our FANTASIA working noise is below marmoset audiogram. Thus, our microscope system can be considered silent or noise-free to marmosets.



During the initial artificial dura implantation Inner diameter is 6.5 mm	Artificial dura window 3 days after the initial implantation Inner diameter is 6.5 mm	Artificial dura window 71 days after the initial implantation Inner diameter is 6.5 mm

Figure 5.3 Artificial dura window in marmosets

A silicone based artificial dura was implanted over marmoset auditory cortex. Vasculature patterns were generally consistent across imaged acquired by 0, 3, 71 days after the initial implantation. The healthy condition of the cortex can be maintained for more than 70 days.

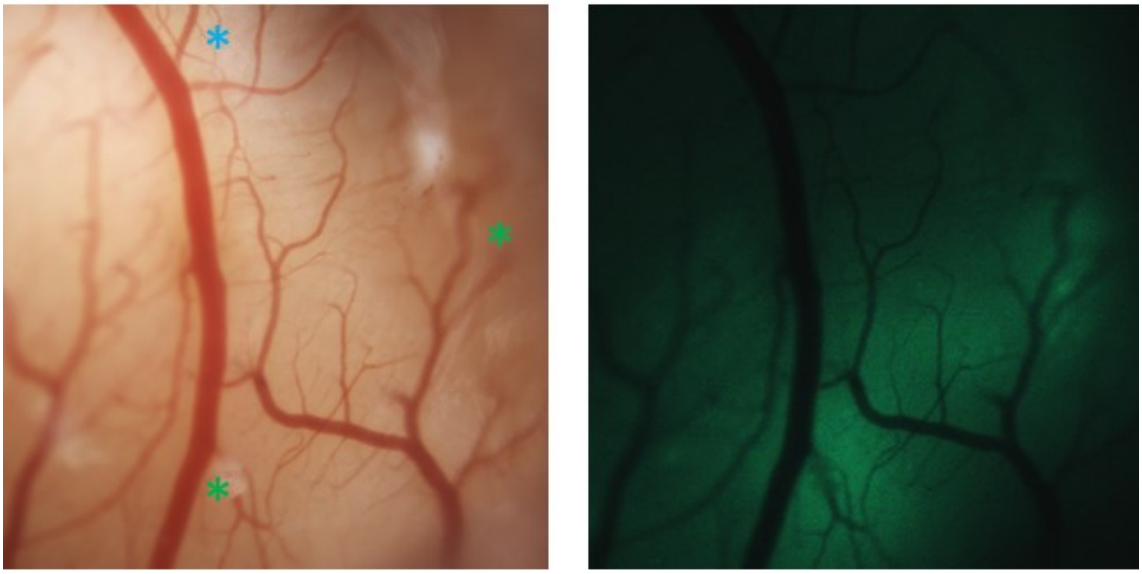


Figure 5.4 Virus injection and labeling in marmosets

Adeno-associated virus (AAV) carrying GCaMP6s and eGFP were injected into marmoset auditory cortex. eGFP expression can be seen through wide-field imaging weeks after the initial injection.

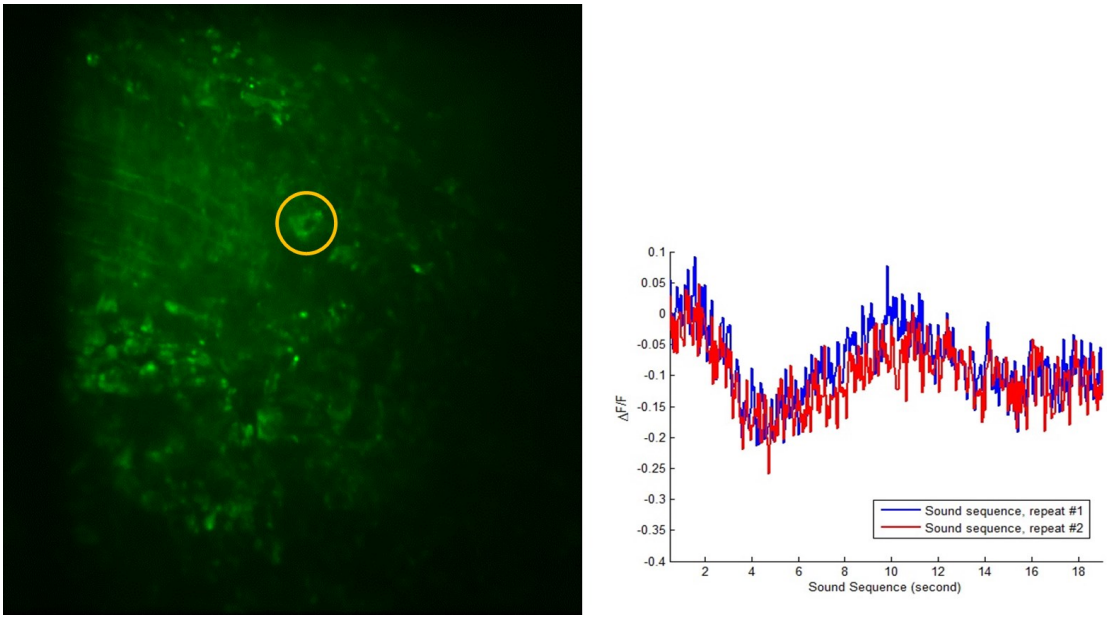


Figure 5.5 Functional fluorescence traces of an exemplar cell recorded from marmoset auditory cortex

An exemplar cell imaged from two-photon imaging at video rate in marmoset auditory cortex. When the same sound sequence played for twice, the fluorescence traces were generally repeatable, and thus are functional.

6. REFERENCES

Abel C, Kössl M (2009) Sensitive Response to Low-Frequency Cochlear Distortion Products in the Auditory Midbrain. *J Neurophysiol* 101(3): 1560–1574,

Agamaite JA, Chang C-J, Osmanski MS, Wang X (2015) A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *J Acoust Soc Am* 138(5): 2906–2928.

Alves-Pinto A, Sollini J, Wells T, Sumner CJ (2016) Behavioural estimates of auditory filter widths in ferrets using notched-noise maskers. *J Acoust Soc Am* 139(2): EL19–EL24.

ANSI (1994) American National Standard Acoustical Terminology. NewYork: American National Standards Institute.

ASA (1960) Acoustical Terminology SI, 1-1960. New York: American Standards Association.

Bandyopadhyay S, Shamma SA, Kanold PO (2010) Dichotomy of functional organization in the mouse auditory cortex. *Nat Neurosci* 13(3): 361–368.

Barnstedt O, Keating P, Weissenberger Y, King AJ, Dahmen JC (2015) Functional microarchitecture of the mouse dorsal inferior colliculus revealed through in vivo two-photon calcium imaging. *J Neurosci* 35(31): 10927–10939.

Bartlett EL, Sadagopan S, Wang X (2011) Fine frequency tuning in monkey auditory cortex and thalamus. *J Neurophysiol* 106(2): 849–859.

Bathellier B, Ushakova L, Rumpel S (2012) Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* 76(2): 435–449.

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436(25): 1161–1165.

Bendor D, Wang X (2010) Neural coding of periodicity in marmoset auditory cortex. *J Neurophysiol* 103(4): 1809–1822.

Bendor D, Osmanski MS, Wang X (2012) Dual pitch processing mechanisms in primate auditory cortex. *J Neurosci* 32(46): 16149–16161.

Bergevin C, McDermott JH, Roy S, Li F, Shera CA, Wang X (2011) Stimulus-frequency otoacoustic emissions as a probe of cochlear tuning in the common marmoset. *Association for Research in Otolaryngology, 34th annual midwinter meeting* poster 371.

Bernstein JGW, Oxenham AJ (2006) The relationship between frequency selectivity and pitch discrimination: effects of stimulus level. *J Acoust Soc Am* 120(6): 3916–3928.

Bezerra BM, Souto A (2008) Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int J Primatol* 29(3): 671–701.

Biebel UW, Langner G (2002) Evidence for interactions across frequency channels in the inferior colliculus of awake chinchilla. *Hear Res* 169(1-2): 151–168.

Bregman MR, Patel AD, Gentner TQ (2016) Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc Natl Acad Sci USA* 113(6): 1666–1671.

Butler RB, Diamond IT, Neff WD (1957) Role of auditory cortex in discrimination of changes of frequency. *J Neurophysiol* 20(1): 108–120.

Capps MJ, Ades HW (1968) Auditory frequency discrimination after transfection of the olivocochlear bundle in squirrel monkey. *Exp Neurol* 21(2): 147–158.

Cedolin L, Delgutte B (2005) Pitch of complex tones: rate–place and interspike interval representations in the auditory nerve. *J Neurophysiol* 94(1): 347–362.

Chen X, Leischner U, Rochefort NL, Nelken I, Konnerth A (2011) Functional mapping of single spines in cortical neurons in vivo. *Nature* 475(7357): 501–505.

Chen T-W, et al. (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* 499(7458): 295–300.

Chung DY, Colavita FB (1976) Periodicity pitch perception and its upper frequency limit in cats. *Percept Psychophys* 20(6): 433–437

Clark WW, Bohne BA (1986) Cochlear damage: Audiometric correlates. *Sensorineural Hearing Loss: Mechanisms, Diagnosis, and Treatment*, eds Collins MJ, Glattke TJ, Harker LA. (University of Iowa Press, Iowa City).

Cynx J, Shapiro M (1986) Percpetion of missing fundamental by a species of songbird (*Sturnus Vulgaris*). *J Comp Psychol* 100(4): 356–360.

Deneux T, Kempf A, Daret A, Ponsot E, Bathellier B (2016) Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nat Commun* 7:12682.

Elliott D, Stein L, Harrison M (1960) Determination of absolute intensity thresholds and frequency difference thresholds in cats. *J Acoust Soc Am* 32(3): 380–384.

Epple G (1968) Comparative studies on vocalization in marmoset monkeys (*Hapalidae*). *Folia Primatol (Basel)* 8(1):1–40.

Evans EF, Pratt SR, Spenner H, Cooper NP (1992) Comparisons of physiological and behavioural properties: auditory frequency selectivity. *Auditory physiology and perception*, eds Cazals Y, Demany L, Horne K (Pergamon, Oxford), pp 159–170.

Fastl H, Weinberger M (1981) Frequency discrimination for pure and complex tones. *Acustica* 49(1): 77–78.

Faulkner A (1985) Pitch discrimination of harmonic complex signals: residue pitch or multiple component discriminations? *J Acoust Soc Am* 78(6): 1993–2004.

Faulstich M, Kossl M (1999) Neuronal response to cochlear distortion products in the anteroventral cochlear nucleus of the gerbil. *J Acoust Soc Am* 105(1): 491–502.

Fay RR (2005) Perception of pitch by goldfish. *Hear Res* 205(1-2): 7–20.

Feng (2013) Spectral integration and neural representation of harmonic complex tones in primate auditory cortex. *Dissertation*.

Gescheider GA (1985) *Psychophysics: method, theory, and application, ed 2.* (Lawrence Erlbaum, New York).

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hear Res* 47(1-2): 103–138.

Goldstein JL (1967) Auditory nonlinearity. *J Acoust Soc Am* 41(3): 676–689.

Goldstein JL (1973) An optimum processor theory for the central formation of the pitch of complex tones. *J Acoust Soc Am* 54(6): 1496–1516.

Green DM, Swets JA (1966) *Signal detection theory and psychophysics* (Robert E. Krieger Publishing Co., Huntington, NY)

Grewe BF, Helmchen F (2009) Optical probing of neuronal ensemble activity. *Curr Opin Neurobiol* 19(5): 520–529.

Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F (2010) High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nat Methods* 7(5): 399–405.

Grimm D, et al. (2008) In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* 82(12): 5887–5911.

Heffner R, Heffner H, Masterton RB (1971) Behavioral measurement of absolute and frequency-difference thresholds in guinea pig. *J Acoust Soc Am* 49(6B): 1888-1895.

Heffner H, Whitfield IC (1976) Perception of the missing fundamental by cats. *J Acoust Soc Am* 59(4): 915–919.

Helmholtz HLF (1863) *Die Lehre von den tonempfindungen, als physiologische grundlage für die theorie der musik* (Braunschweig, F. Vieweg und Sohn).

Henning GB, Grosberg SL (1968) Effect of harmonic components on frequency discrimination. *J Acoust Soc Am* 44(5): 1386–1389.

Houtsma AJM, Smurzynski J (1990) Pitch identification and discrimination for complex tones with many harmonics. *J Acoust Soc Am* 87(1): 304–310.

ISO (1975) *ISO16:1975 - Acoustics - Standard Tuning Frequency (Standard Musical Pitch)* (International Organization for Standardization, Geneva).

Issa JB, et al. (2014) Multiscale optical Ca2+ imaging of tonal organization in mouse auditory cortex. *Neuron* 83(4): 944–959.

Joly O, et al. (2014) A perceptual pitch boundary in a non-human primate. *Front Psychol* 5(September): 998.

Joris PX et al (2011) Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *Proc Natl Acad Sci USA* 108(42): 17516–17520.

Kaernbach C, Bering C (2001) Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *J Acoust Soc Am* 110(2): 1039–1048.

Kato HK, Gillet SN, Isaacson JS (2015) Flexible sensory representations in auditory cortex driven by behavioral relevance. *Neuron* 88(5): 1027–1039.

Klinge A, Itatani N, Klump GM (2010) A comparative view on the perception of mistuning: constraints of the auditory periphery. *The neurophysiological basis of auditory perception*, eds Lopez-Poveda EA, Palmer AR, Meddis R (Springer, New York), pp 465–475.

Kohlrausch A, Sander A (1995) Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *J Acoust Soc Am* 97(3): 1817–1829.

Kojima S (1990) Comparison of auditory functions in the chimpanzee and human. *Folia Primatol* 55(2): 62–72.

Langner G (1997) Neural processing and representation of periodicity pitch. *Acta Otolaryngol Suppl* 532: 68–76.

Letzkus JJ, Wolff SBE, Meyer EMM, Tovote P, Courtin J, Herry C, Lüthi A (2011) A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* 480(7377): 331–335.

Licklider JCR (1956) Auditory frequency analysis. *Information Theory*, eds Cherry C (Academic Press, New York), pp 253–268.

Long G, Clark W (1984) Detection of frequency and rate modulation by the chinchilla. *J Acoust Soc Am* 75(4): 1184–1190.

Lu T, Liang L, Wang X (2001) Neural representation of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophys* 85(6): 2364–2380.

Maor I, Shalev A, Mizrahi A (2016) Distinct spatiotemporal response properties of excitatory versus inhibitory neurons in the mouse auditory cortex. *Cereb Cortex* 26(11): 4242–4252.

May BJ, Kimar S, Prosen CA (2006) Auditory filter shapes of CBA/CaJ mice: behavioral assessments. *J Acoust Soc Am* 120(1): 321–330.

McAlpine D (2004) Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. *J Neurophysiol* 92(3): 1295–1311.

McDermott JH, Oxenham AJ (2008) Music perception, pitch, and the auditory system. *Curr Opin Neurobiol* 18(4):452–463.

Meddis R, O'Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102(3): 1811–1820.

Meddis R, O'Mard L (2006) Virtual pitch in a computational physiological model. *J Acoust Soc Am* 102(6): 3861–3869.

Michelet P, McLaughlin M, van der Heijden M, Joris P (2011) Synchronization to pure and amplitude-modulated tones in the auditory nerve of macaque monkey. *Association for Research in Otolaryngology, 34th annual midwinter meeting* poster 666.

Micheyl C, Divis K, Wrobleski DM, Oxenham AJ (2010) Does fundamental-frequency discrimination measure virtual pitch discrimination? *J Acoust Soc Am* 128(4): 1930–1942.

Micheyl C, Ryan CM, Oxenham AJ (2012) Further evidence that fundamental-frequency difference limens measure pitch discrimination. *J Acoust Soc Am* 131(5): 3989–4001.

Miller CT, et al. (2016) Marmosets: a neuroscientific model of human social behavior. *Neuron* 90(2): 219–233.

Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J Acoust Soc Am* 74(3): 750–753.

Moore BCJ, Glasberg BR (1990) Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J Acoust Soc Am* 87(5): 2163–2177.

Moore BCJ, Moore GA (2003) Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hear Res* 182(1-2): 153–163.

Moore BCJ (2012) *An Introduction to the Psychology of Hearing, ed 6*. (Emerald, Bingley, UK), pp 86.

Nathanson JL, Yanagawa Y, Obata K, Callaway EM (2009) Preferential labeling of inhibitory and excitatory cortical neurons by endogenous tropism of adeno-associated virus and lentivirus vectors. *Neuroscience* 161(2): 441–450.

Nelson DA, Kiester TE (1978) Frequency discrimination in the chinchilla. *J Acoust Soc Am* 64(1): 114–126.

Niemiec AJ, Yost WA, Shofner WP (1992) Behavioral measures of frequency selectivity in the chinchilla. *J Acoust Soc Am* 92(5): 2636–2649.

Norman-Haignere S, Kanwisher N, McDermott JH (2013) Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J Neurosci* 33(50): 19451–19469.

Ohm GS (1843) Ueber die definition des tones, nebst daran geknüpfter theorie der sirene und ähnlicher tonbildender vorrichtungen. *Ann Phys Chem* 59: 513–565.

Okanoya K (2000) Perception of missing fundamental in zebra finches and Bengalese finches. *J Acoust Soc Jpn (E)*. 21(2): 63-68.

Osmanski MS, Wang X (2011) Measurement of absolute auditory thresholds in the common marmoset (*Callithrix jacchus*). *Hear Res* 277(1-2): 127–133.

Osmanski MS, Song X, Wang X (2013) The role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (*Callithrix jacchus*). *J Neurosci* 33(21): 9161–9168.

Osmanski MS, Song X, Guo Y, Wang X (2016) Frequency discrimination in the common marmoset (*Callithrix jacchus*). *Hear Res* 341: 1–8.

Oxenham AJ, Micheyl C (2013) Pitch perception: dissociating frequency from fundamental-frequency discrimination. *Adv Exp Med Biol* 787: 137–145.

Oxenham AJ, Micheyl C, Keebler MV (2009) Can temporal fine structure represent the fundamental frequency of unresolved harmonics? *J Acoust Soc Am* 125(4): 2189–2199.

Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S (2011) Pitch perception beyond the traditional existence region of pitch. *Proc Natl Acad Sci U S A*. 108(18): 7629–7634.

O'Shea DJ, et al. (2016) The need for calcium imaging in nonhuman primates: New motor neuroscience and brain-machine interfaces. *Exp Neurol* 287(Pt 4): 437–451.

Penagos H, Melcher JR, Oxenham AJ (2004) A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging. *J Neurosci* 24(30): 6810–6815.

Peron S, Chen T-W, Svoboda K (2015) Comprehensive imaging of cortical networks. *Curr Opin Neurobiol* 32: 115–123.

Plack CJ, Oxenham AJ, Fay RR, Popper AN (2005) *Pitch: Neural Coding and Perception* (Springer, New York).

Plomp R (1964) The Ear as a Frequency Analyzer. *J Acoust Soc Am* 36(9): 1628–1636.

Plomp R, Mimpen AM (1968) The Ear as a Frequency Analyzer. II. *J Acoust Soc Am* 43(4): 764–767.

Pressnitzer D, Patterson RD (2001) Distortion products and the pitch of harmonic complex tones. *Physiological and Psychophysical Bases of Auditory Function*, eds Breebaart DJ, Houtsma AJM, Kohl-rausch A, Prijs VF, Schoonhoven R (Shaker, Maastricht), pp. 97–104.

Prosen CA, Moody DB, Sommers MS, Stebbins WC (1990) Frequency discrimination in the monkey. *J Acoust Soc Am*. 88(5): 2152–2158.

Recanzone GH, Jenkins WM, Hradek GT, Merzenich MM (1991) A behavioral frequency discrimination paradigm for use in adult primates. *Behav Res Methods Instrum Comput* 23(3): 357–369.

Recio-Spinoso A, Temchin AN, van Dijk P, Fan YH, Ruggero MA (2005) Wiener-kernel analysis of responses to noise of chinchilla. *J Neurophysiol* 93(6): 3615–3634.

Remington ED, Osmanski MS, Wang X (2012) An operant conditioning method for studying auditory behaviors in marmoset monkeys. *PLoS One* 7(10): e47895.

Roe AW (2007) Long-term optical imaging of intrinsic signals in anesthetized and awake monkeys. *Appl Opt* 46(10): 1872–1880.

Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. *Nat Neurosci* 13(3): 353–60.

Rothschild G, Cohen L, Mizrahi A, Nelken I (2013) Elevated correlations in neuronal ensembles of mouse auditory cortex following parturition. *J Neurosci* 33(31): 12851–12861.

Ruiz O, et al. (2013) Optogenetics through windows on the brain in the nonhuman primate. *J Neurophysiol* 110(6): 1455–1467.

Sadakane O, et al. (2015) Long-term two-photon calcium imaging of neuronal populations with subcellular resolution in adult non-human primates. *Cell Rep* 13(9): 1989–1999.

Schouten JF (1938) The perception of subjective tones. *Proc K Ned Akad Wet* 41: 1086–1093.

Schouten JF (1940) The residue and the mechanism of hearing. *Proc K Ned Akad Wet* 43: 991–999.

Schroeder M (1970) Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans Inf Theory* 16(1): 85–89.

Schulze H, Hess A, Ohl FW, and Scheich H (2002) Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil. *Eur J Neurosci* 15(2): 1077–1084.

Schulze H, Langner G (1997) Representation of periodicity pitch in the primary auditory cortex of the Mongolian gerbil. *Acta Otolaryngol Suppl* 532: 89–95.

Schulze H, Langner G (1999) Auditory cortical responses to amplitude modulations with spectra above frequency receptive fields: evidence for wide spectral integration. *J Comp Physiol [A]* 185(6): 493–508.

Seebeck A (1841) Beobachtungen über einige bedingungen der entstehung von tönen. *Ann Phys Chem* 53(7):417–436.

Seidemann E, et al. (2016) Calcium imaging with genetically encoded indicators in behaving primates. *Elife* 5(2016JULY): 1–19.

Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am* 95(6): 3529–3540.

Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J Acoust Soc Am* 107(5): 2631–2644.

Shera CA, Guinan JJ (2003) Stimulus-frequency-emission group delay: a test of coherent reflection filtering and a window on cochlear tuning. *J Acoust Soc Am* 113(5): 2762–2772

Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99(5): 3318–3323.

Shera CA, Guinan JJ, Oxenham AJ (2010) Otoacoustic estimation of cochlear tuning: validation in the chinchilla. *J Assoc Res Otolaryngol* 11(3): 343–365.

Shofner WP (2000) Comparison of frequency discrimination thresholds for complex and single tones in chinchillas. *Hear Res* 149(1-2): 106–114.

Shofner W (2011) Perception of the missing fundamental by chinchillas in the presence of low-pass masking noise. *J Assoc Res Otolaryngol* 12(1): 101–112.

Shofner WP, Chaney M (2013) Processing pitch in a nonhuman mammal (*Chinchilla laniger*). *J Comp Psychol* 127(2): 142–153.

Siegel JH, Cerka AJ, Recio-Spinoso A, Temchin AN, van Dijk P, Ruggero MA (2005) Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering. *J Acoust Soc Am* 118(2): 2434–2443.

Sinnott JM, Petersen MR, Hopp SL (1985) Frequency and intensity discrimination in humans and monkeys. *J Acoust Soc Am*. 78(6): 1977–1985.

Sinnott JM, Owren MJ, Petersen MR (1987) Auditory frequency discrimination in primates: Species differences (*Cercopithecus, Macaca, Homo*). *J Comp Psychol* 101(2): 126–131.

Sinnott JM, Brown CH, Brown FE (1992) Frequency and intensity discrimination in Mongolian gerbils, African monkeys and humans. *Hear Res* 59(2): 205–212.

Smoorenburg GF, Gibson MM, Kitzes LM, Rose JE, Hind JE (1976) Correlates of combination tones observed in the response of neurons in the anteroventral cochlear nucleus of the cat. *J Acoust Soc Am* 59(4): 945–962.

Slaney M (1998) "Auditory Toolbox Version 2", Technical Report #1998-010, Interval Research Corporation. http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/

Spiegel MF, Watson CS (1984) Performance on frequency-discrimination tasks by musicians and nonmusicians. *J Acoust Soc Am* 76(6): 1690–1695.

Stebbins WC (1973) Hearing in old world monkeys (*Cercopithecinae*). *Am J Phys Anthropol* 38(2): 357–364.

Terhardt E (1974) Pitch, consonance, and harmony. *J Acoust Soc Am*. 55(5): 1061–1069.

Thurlow WR, Small AM (1955) Pitch perception for certain periodic auditory stimuli. *J Acoust Soc Am* 27(1): 132–137.

Tomlinson RWD, Schwarz DWF (1988) Perception of the missing fundamental in nonhuman primates. *J Acoust Soc Am* 84(2): 560–565.

Tsuji J, Liberman MC (1997) Intracellular labeling of auditory nerve fibers in guinea pig: central and peripheral projections. *J Comp Neurol* 381(2): 188–202.

Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435(7040): 341–346.

Watakabe A, et al. (2015) Comparative analyses of adeno-associated viral vector serotypes 1, 2, 5, 8 and 9 in marmoset, mouse and macaque cerebral cortex. *Neurosci Res* 93: 144–157.

Wienicke A, Häusler U, Jürgens U (2001) Auditory frequency discrimination in the squirrel monkey. *J Comp Physiol A* 187(3): 189–195.

Wier C, Jesteadt W, Green D (1977) Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am* 61(1):178–184.

Worley KC et al (2014) The common marmoset genome provides insight into primate biology and evolution. *Nat Genet* 46(8): 850–857.

Zeitlin LR (1964) Frequency Discrimination of Pure and Complex Tones. *J Acoust Soc Am* 36(5): 1027.

7. CURRICULUM VITAE

Xindong Song **November 7, 2016**

Education History

PhD expected	2016	Biomedical Engineering, Johns Hopkins University Mentor: Dr. Xiaoqin Wang
M.S	2008	Biology, Tsinghua University
B.E.	2004	Electronic Engineering, Tsinghua University

Academic Awards

- Johns Hopkins School of Medicine, the 39th Young Investigators' Day, Bae Gyo Jung Research Award, 2016

Publications

- Osmanski MS*, **Song X***, GuoY, Wang X (2016) "Frequency discrimination in the common marmoset (*Callithrix jacchus*)". *Hearing Research*. 341:1-8 **(*contributes equally, cover story**)
- **Song X**, Osmanski MS, Guo Y, Wang X (2016) "Complex pitch perception mechanisms are shared by humans and a New World monkey". *Proceedings of the National Academy of Sciences of the USA*. 113:781-786 (**cover story**)
- Osmanski MS, **Song X**, Wang X (2013) "The role of harmonic resolvability in pitch perception in a vocal nonhuman primate, the common marmoset (*Callithrix jacchus*)". *The Journal of Neuroscience* 33: 9161–9168.
- Pan L*, **Song X***, Xiang G, Wong A, Xing W, Cheng J (2009). "First-spike rank order as a reliable indicator of burst initiation and its relation with early-to-fire neurons". *IEEE Transactions on Biomedical Engineering.* 56:1673-1682. (***contributes equally**)

- Pan L*, **Song X***, Xiang G, Zhu J, Cheng J (2009). "Effects of disinhibition on spatiotemporal pattern of neuronal first recruitment in neuronal networks". *Progress in Natural Science* 19:615-621. (***contributes equally**)

- **Song X** (2008). "Fusion of protein structure database and nucleotide sequence database for exploration of the synonymous codons' functions". *Journal of Information and Computational Science.* 5:1437-1443.

- Xiang G, Pan L, Huang L, Yu Z, **Song X**, Cheng J, Xing L, Zhou Y (2007). "Microelectrode array-based system for neuropharmacological applications with cortical neurons cultured in vitro". *Biosensors and Bioelectronics*. 22:2478-2484.

Oral Presentations

- "Pitch Perception in Marmosets". In nano-symposium "marmoset neurobiology" of Society for Neuroscience, 46th annual meeting, 2016

- "Towards Understanding Functional Neuronal Networks Underlying Pitch Perception". Neurocog Collective, Nosara, Costa Rica, 2016

- "How to Define Frequency Resolution in Audition" Auditory Splash Workshop, MIT, Boston, MA, USA, 2016

- "Pitch Perception", invited lecture in the course "Psychoacoustics" (PY 550.517), Peabody Conservatory, Baltimore, MD, USA, 2015

- "Pitch Perception in Marmosets", invited talk in the Primate Neuroscience Workshop, Tsinghua University, Beijing, China, 2015

Posters and abstracts

- **Song X**, Guo Y, Li X, Wang X (2015) "Flexible, nimble, and quiet two-photon microscope platform for auditory functional imaging of awake marmosets". *Society for Neuroscience, 45th annual meeting*. 732.15.

- Guo Y, Osmanski MS, **Song X**, Wang X (2015) "Measurement of acoustic frequency discrimination thresholds in common marmosets (Callithrix jacchus)". *Society for Neuroscience, 45th annual meeting*. 508.06.

- **Song X**, Osmanski MS, Guo Y, Wang X (2015) "Possible origins or human-like pitch perception mechanisms". *Association for Research in Otolaryngology, 38th annual midwinter meeting*. PS-118.

- **Song X**, Osmanski MS, Guo Y, Wang X (2014) "Possible origins or human-like pitch perception mechanisms". *Society for Neuroscience, 44th annual meeting*. 181.27.
- **Song X**, Osmanski MS, Wang X (2013) "Measurement of pitch discrimination thresholds in the common marmoset (Callithrix jacchus)". *Society for Neuroscience, 43th annual meeting*. 354.08.
- Osmanski MS, **Song X**, Wang X (2013) "Defining harmonic resolvability in the common marmoset (Callithrix jacchus)". *Society for Neuroscience, 43th annual meeting*. 354.07.
- Bian C*****, **Song X***, Liu Z, Zhang H (2005). "Design proposal of imaging activities of cultured neural network on a silicon substrate with neural-electronic-optical integrated microsystem". *27th Annual International Conference of the Engineering in Medicine and Biology Society, (IEEE-EMBS 2005)*. 7600-7603. (oral presentation, ***contributes equally**)